DOCUMENT RESUME

ED 144 944 95 TM 005 577

AUTHOR Bryk, Anthony S.; Weisberg, Herbert I.
TITLE A Developmental Perspective on Statistical Methods of Removing Bias in Quasi-Experiments. Final Report.
INSTITUTION Contract Research Corp., Belmont, Mass.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Sep 75
CONTRACT NIE-C-74-0125
NOTE 107p.; Best copy available

EDRS PRICE MF-$0.83 HC-$6.01 Plus Postage.
DESCRIPTORS *Computer Programs; Educational Experiments; Individual Development; Intervention; *Mathematical Models; Personal Growth; Program Evaluation; Reliability; *Research Design; Research Problems; Simulation; *Statistical Analysis; *Statistical Bias

ABSTRACT

A mathematical model was developed to represent the way data are generated in quasi-experiments designed to study interventions in growth systems. This model explicitly represents three important processes: (1) the nature of individual growth; (2) the process by which individuals are assigned to treatment groups; and (3) the nature of the treatment effect. The way in which these three factors relate to the effectiveness of various methods is explored via both mathematical analysis and computer simulation. Results suggest that, in general, none of the statistical methods is very effective in removing the biases resulting from non-random assignment. It is concluded that the traditional pre-post, treatment versus control design with non-random assignment is not adequate in growth situations, and that research should focus on the development of experimental designs and corresponding analyses better suited to the dynamic nature of developmental processes. The FORTRAN IV computer program used for the simulation is appended. (Author/MV)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED144444

BEST COPY AVAILABLE

CONTRACT RESEARCH CORPORATION

25 Flanders Road
Belmont
Massachusetts 02178

A Developmental Perspective on Statistical Methods of Removing Bias in Quasi-Experiments

Final Report

September 1975

Submitted to:

Dr. Andrew Porter
National Institute of Education
Washington, D. C.

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM005 577

FINAL REPORT

A Developmental Perspective on Statistical Methods for Removing Bias in Quasi-Experiments

National Institute of Education Project No. C-74-0125

Project Director: Herbert I. Weisberg

Contract Research Corporation

1975

Authors: Anthony S. Bryk
Herbert I. Weisberg

The research reported herein was performed pursuant to a grant contract with the National Institute of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official National Institute of Education position or policy.

Table of Contents

	Page
Abstract	
1.0 Introduction	1
2.0 Analysis Strategies	5
3.0 The Model	8
3.1 Natural Growth	8
3.2 Treatment Effect	11
3.3 Assignment Process	13
4.0 Analysis of the Linear Growth System	16
4.1 The Linear Model	16
4.2 Properties of the Linear System	17
4.3 Linear Adjustment Strategies under the Linear Growth Model	19
5.0 Computer Simulation	25
5.1 The Simulation Program	25
5.2 Growth Models Considered	27
5.3 Simulation Results	28 31
6.0 Summary and Conclusions	31
Bibliography	
Appendix A - Tables 1 through 16	
Appendix B - User's Guide to the Computer Program	
Appendix C - Fortran IV Computer Program for Generating Quasi-Experimental Data in Situations Involving Growth and Examining the Effectiveness of Analysis Strategies	

Abstract

A Developmental Perspective on Statistical Methods for Removing Bias in Quasi-Experiments

Anthony S. Bryk
Herbert I. Weisberg

Randomized experiments are often not feasible in educational research. To analyze comparative studies where random assignment to treatments has not been employed, a number of "adjustment" strategies have been suggested. There is much controversy about the adequacy of such strategies.

In many educational and social settings, an experimental program is a dynamic intervention in an ongoing developmental process. To understand the effects of using various statistical adjustment methods, one must consider the nature of such processes in some detail. In this report we develop a mathematical model to represent somewhat realistically the way data are generated in quasi-experiments designed to study interventions in growth systems. This model explicitly represents three important processes: (1) the nature of individual growth; (2) the process by which individuals are assigned to treatment groups; and (3) the nature of the treatment effect.

The way in which these three factors relate to the effectiveness of various methods is explored in this report via both mathematical analysis and computer simulation. Results suggest that, in general, none of the statistical methods is very effective

in removing the biases resulting from non-random assignment. We conclude that the traditional pre-post, treatment versus control design with non-random assignment is not adequate in growth situations, and that research should focus on the development of experimental designs and corresponding analyses better suited to the dynamic nature of developmental processes.

1.0 Introduction

In recent years the federal government has mounted several large-scale evaluations of the effectiveness of various educational programs. For a variety of practical, ethical, and political reasons, strict adherence to the canons of experimental research design has generally been impossible. Our interest in the general problem of analyzing non-experimental data has grown out of an involvement in two major evaluations: Head Start Planned Variation and Follow Through. Both of these studies were quasi-experiments involving the comparison of several curricular models, but with no random assignment of subjects to treatments.

In the last few years there has been much controversy over how to draw valid inferences from such quasi-experiments, and, indeed, whether it is even possible to do so. Tucker, Damarin, and Messick (1966) have suggested residual analysis. Cronbach and Furby (1970) and Werts and Linn (1970) have argued that the General Linear Model and, in particular, analysis of covariance, is the only appropriate approach. More recently, Smith (1972) and Kenny (1975) have suggested that gain scores may in fact be the answer. In addition, several other approaches have been suggested in this context, including direct and indirect standardization (Wiley, 1971), and matching (Rubin, 1973). We too have contributed to the confusion with our value-added analysis (Bryk and Weisberg, 1974, 1976).

On the other hand such writers as Lord (1967) and Campbell and Erlebacher (1970) have pointed to problems in the use of such adjustment strategies. These problems have led some (c.f. Gilbert, Light and Mosteller [1975] and Riecken and Boruch [1974]) to push hard for the implementation of randomization if at all possible.

In trying to understand the extent to which various analysis strategies could provide unbiased estimates of treatment effects, we became frustrated by the absence in the literature of realistic models which focus on the process by which the data are generated. Some educational researchers seem to have valuable intuition about the processes generating quasi-experimental data, but they do not express these ideas in the form of a coherent model. Most mathematically-oriented methodologists, on the other hand, seem to accept uncritically the general linear model which has been so useful in many non-educational settings.

In thinking about quasi-experimental data, we gradually reached the conclusion that three main factors must be considered: (1) the nature of individual growth; (2) the process by which individuals are assigned to treatment groups; and (3) the nature of the treatment effect.

In education, as in many other domains, we are typically examining systems which are fundamentally dynamic and adaptive. Individuals are growing, and changing in interaction with one another and the environment even in the absence of any external intervention. Such phenomena are not explicitly considered in the linear model, but are rather treated as unmeasured nuisance variables.

Explicit consideration of the process by which individuals are assigned to groups is almost non-existent in the literature (Kenny, 1975 is an exception). Most analysts begin by assuming the convenient fiction that there exist two distinct populations from which the program and control groups are randomly selected. Each population is characterized by a multivariate (usually normal) distribution on the variables of interest. The bias reduction properties of various strategies are examined in terms of the parameters of these two distributions (see for example, Cochran and Rubin, 1974).

An alternative way to conceptualize this situation is to imagine that sampling is from a single population of individuals who are then assigned non-randomly to a program and control group. We prefer this latter fiction because it allows us to focus explicitly on the mechanism by which subjects are assigned to groups.

The treatment effect is traditionally modeled as a constant increment for all subjects, although consideration of "aptitude-treatment interactions" permits some flexibility. We propose to consider models in which the treatment effect may be a function of the individual's own status or even the group context in which he or she is developing.

Our purpose in this paper is to assess the ability of various data analysis strategies to remove biases which result from nonrandom assignment of subjects to groups. To do this, we examine each technique in terms of a mathematical model which

represents somewhat realistically the way quasi-experimental data arise. The remainder of this paper consists of five sections.

First, we discuss briefly the various analysis strategies which have been suggested for analyzing educational quasi-experiments. A set of sixteen strategies whose properties are to be explored is presented.

In section 3 we present in detail the mathematical model upon which this investigation is based. Explicitly represented in this general model are specific models of the three main processes mentioned above: individual growth, assignment of subjects to groups, and the treatment effect.

In section 4 we develop some analytical results under the assumption of a relatively simple model. Some of these results relate to growth systems in general and may have broad application in studying developmental processes. Others relate more specifically to the problem of eliminating bias in the analysis of quasi-experimental data.

In section 5 we discuss the computer simulation program based on the mathematical model described in section 3. This program has been used to explore the properties of strategies in more complex situations. Results of some simulation runs are presented.

Finally, in section 6, we discuss the implications of our research for evaluation practice. Suggestions for the development of evaluation methodology along certain lines are made and possible future extensions of the present work discussed.

2.0 Analysis Strategies

In this section we discuss briefly the set of analysis strategies whose properties are to be studied. For simplicity, we consider only the situation where one treatment group is being compared to one control group. We also assume that the information available to us consists of the pre-test, the post-test, and the age for each individual. Some methods make use of other "covariates" besides pre-test and age. We are including these two because (1) they are generally quite important in educational contexts and (2) the inclusion of other variables would add tremendous complexity.

Most techniques for analyzing quasi-experimental data fall under the general heading of linear adjustments. These adjustments have the following general form:

$$\hat{\alpha} = \bar{Y}_{2p} - \bar{Y}_{2c} - \beta\,(\bar{Y}_{1p} - \bar{Y}_{1c}) - \gamma(\bar{a}_{1p} - \bar{a}_{1c})$$

where

α = true treatment effect

$\bar{Y}_{2p}$ ($\bar{Y}_{2c}$) = post-test mean for program (control) group

$\bar{Y}_{1p}$ ($\bar{Y}_{1c}$) = pre-test mean for program (control) group

$\bar{a}_{1p}$ ($\bar{a}_{1c}$) = mean age at pre-test time for program (control) group

The coefficients β and γ characterize the particular linear adjustment method. For example, if $\beta = 0$ and $\gamma = 0$, our estimate of α is the simple difference of post-test means; while if $\beta = 1$ and $\gamma = 0$, we have gain scores. In all, thirteen such linear adjustment strategies have been listed in Table 1.

In addition to the linear adjustments, we also consider two forms of adjustment by subclassification (Cochran, 1968), which is a particular type of standardization (Wiley, 1971). The basic idea is to estimate the post-test means the treatment groups would have had if the pre-test distribution of the treatment group were the same as that of the controls. To approximate this, class boundaries are found which divide the pre-test distribution for the control group into six approximately equal size classes (in terms of frequency). Subjects in all groups are assigned to a class on the basis of these pre-test boundaries. Post-test means are computed for each pre-test group. These six means are then proportionally weighted by the number in the cell to form an overall mean which is compared with the control group mean. The difference between the two is our subclassification estimate of the treatment effect.

The second subclassification technique is based on both the pre-test and age. We have elected to divide each variable into three classes, resulting in a nine-cell design.

Another strategy considered is a variant of the value-added approach which we have developed recently (Bryk and Weisberg, 1974). This method does not utilize the control group information per se. Rather than using the relationship between pre-test and post-test, the value-added technique utilizes the individual's pre-test status and age to project a post-test status in the absence of intervention. The actual post-test is then compared to the projected post-test. The average difference across subjects within the treatment group is an estimate of the

treatment effect. In its application here, we use the ordinary least squares regression of the pre-test on age as the basis for our projections.

Another approach to analysis which has received some attention is matching (c.f. Rubin, 1974). Matching is, however, most effective as a component of the design rather than the analysis. Post hoc matching is possible, but its effectiveness generally depends critically on having a control group much larger than the treatment group, so that good matches can be found. These factors, we felt, would make a fair comparison of matching with the other analysis strategies mentioned above very difficult.

3.0 The Model

In this section we present a general mathematical model to represent the way in which quasi-experimental data arise in many educational settings involving growth. As in most models of dynamic processes, we shall employ the differential equation as our basic conceptual tool. As mentioned above, the model has three main components: representing the processes of natural growth, the effect of a treatment, and the assignment of individuals to treatment groups.

3.1 Natural Growth

We assume that each individual is characterized by a growth rate at any point in time which would apply in the absence of the experimental program. This growth rate may in general be a function of one or more parameters and of the individual's current status. Let $G_i(t)$ represent individual i's natural growth function in terms of some dimension of interest. Moreover, let $\underset{\sim}{\pi}_i = (\pi_{i1}, \ldots \pi_{ik})$ represent a vector of parameters characterizing individual i's growth rate. Then G can be defined as that function which satisfies a differential equation of the form

$$\frac{\partial G_i}{\partial t} = g_i\ (\underset{\sim}{\pi}_i,\ G_i(t)) \qquad (3.1)$$

Of course there are many possible candidates for g_i. We will present four which seem particularly useful. The first two are selected because of their simplicity. The latter two suggest themselves because they can model a broad range of non-linear growth, and have seen some use in modeling certain developmental

phenomena, such as human physical growth (e.g., Bock et al, 1973).

These are given by

$$\frac{\partial G_i}{\partial t} = \pi_{1i} \tag{3.2}$$

$$\frac{\partial G_i}{\partial t} = \pi_{1i} + \pi_{2i} t \tag{3.3}$$

$$\frac{\partial G_i}{\partial t} = \pi_{1i}(1 - G_i) \tag{3.4}$$

$$\frac{\partial G_i}{dt} = \pi_{1i} G_i(1 - G_i) \tag{3.5}$$

In analyzing the first three of these models, it is convenient to define t_{oi} as the onset time for individual i. That is we assume

$$g(\underset{\sim}{\pi}_i, G_i(t)) = 0 \quad \text{for } t < t_{oi} \tag{3.6}$$

and that g has one of the above forms for $t > t_{oi}$. For the fourth case, it is convenient to define $t_{.5i}$ as the time at which half of "full" growth is reached. With this notation we can express the natural growth curves as

$$G_i(t) = \pi_{1i}(t-t_{oi}) \quad \text{(linear)} \tag{3.7}$$

$$G_i(t) = \pi_{1i}(t-t_{oi}) + \pi_{2i}(t-t_{oi})^2 \quad \text{(quadratic)} \tag{3.8}$$

$$G_i(t) = 1-e^{-\pi_{1i}(t-t_{oi})} \quad \text{(exponential)} \tag{3.9}$$

$$G_i(t) = \frac{1}{1+e^{-\pi_{1i}(t-t_{.5i})}} \quad \text{(logistic)} \tag{3.10}$$

So far we have considered the growth for each individual. What about the population? To model the process of selecting a sample from a given population, we propose to consider the π_i's as random variables with theoretically reasonable distributions. The actual choice of specific theoretical distributions for the growth parameters, will depend on the substantive meaning of the parameters themselves.

As a simplifying assumption we shall assume that the various parameters are mutually independent. This assumption greatly facilitates the analysis of growth systems. In future work, we hope to explore more general models and to consider explicitly the realism of the independence assumption.

To model some situations of interest, and because it is often used as a covariate in developmental research, it is useful to have a representation of each individual's age. To accomplish this, we do not generate t_{oi}, the onset time directly, but as the sum of time of birth and onset age. Let

b_i = time of birth for individual i

δ_i = onset age for individual i

Then

$$t_{oi} = b_i + \delta_i \tag{3.11}$$

and

$$a_i(t) = t - b_i = \text{age at time } t \tag{3.12}$$

Finally, to make our model of growth more realistic, we allow the possibility of a random component to the growth. The systematic natural growth discussed so far may be considered as

the smoothed trend underlying an observed time series. The random component reflects in part the instability characterizing our assessment instruments which is commonly termed measurement error. The random component also reflects, however, real characteristics of individuals. While conceptually it may be useful to consider individual growth over a period of time as a smooth curve, i.e., the systematic component, in reality the individual growth also consists of discontinuities, instability, and real variation. Thus, the random component is meant to include measurement error, but is not limited exclusively to it.

Mathematically we represent the random component by

$R_i(t)$ = random component of growth

where

$$E\,(R_i(t)\,|\,G) = 0 \tag{3.13}$$

$$Cov\,(R_i(t_1),\ R_i(t_2)) = 0 \tag{3.14}$$

The observed score for individual i at time t is then represented by

$$Y_i(t) = G_i(t) + R_i(t) \tag{3.15}$$

3.2 Treatment Effect

So far we have developed a model to represent growth of a particular group on a particular dimension. Now suppose that during some specified period of time, a subset of the group is to be exposed to an experimental program. We put off until the next section the question of how this group is to be selected, and focus here on representing the treatment effect for this group.

First let

t_1 = start of the program (time of pre-test)

t_2 = end of program (time of post-test)

We assume that the program operates between times t_1 and t_2, and measurements are obtained at these times.

We conceive of the program as effecting an alteration in the natural growth rate. We define a treatment effect rate,

$$\frac{\partial I_i}{\partial t} = 0 \text{ for } t < t_1 \tag{3.16}$$

$$= h(\underset{\sim}{\varepsilon}_i, G_i(t)) \text{ for } t_1 \leq t \leq t_2$$

$$= 0 \text{ for } t > t_2$$

where $\underset{\sim}{\varepsilon}$ is a vector or parameters characterizing individual i. The treatment effect rate can in principle be a function of one or more parameters as well as the current growth status. By proper choice of the h function, we can model a broad range of educational paradigms.

We can now define a treatment-augmented growth rate,

$$\frac{\partial S_i}{\partial t} = \frac{\partial G_i}{\partial t} + \phi_i \frac{\partial I_i}{\partial t} \tag{3.17}$$

where ϕ_i = 1 if i is in the program group

0 if i is in the control group

and a treatment-augmented growth curve,

$$S_i(t) = \int \frac{\partial S_i}{\partial t} \tag{3.18}$$

We define the observed score at time t by

$$Y_i(t) = S_i(t) + R_i(t) \tag{3.19}$$

and for any individual the treatment effect at time t_2 is given by

$$\alpha_i = S_i(t_2) - G_i(t_2) \tag{3.20}$$

For example, in the simple case of a linear growth system with a constant treatment effect rate,

$$\frac{\partial I_i}{\partial t} = \varepsilon \tag{3.21}$$

$$\frac{\partial S_i}{\partial t} = \pi_i + \varepsilon \tag{3.22}$$

This yields upon integration,

$$S_i(t) = \pi_i(t-t_{oi}) + \varepsilon(t-t_1) \tag{3.23}$$

for $t_1 \leq t \leq t_2$

and a constant treatment effect across the subjects at t_2 of

$$\alpha = \varepsilon(t_2-t_1) \tag{3.24}$$

3.3. Assignment Process

In this section we discuss the model to represent the assignment process. Let us assume that the total sample consists of n individuals, and that, or these, n_p are to be assigned to the program group and n_c to the control group ($n_c + n_p = n$).

We want to be able to represent a general situation where each individual's probability of being assigned to the program group may be random or may to varying degrees depend on his growth parameter values. Moreover, rather than basing selection directly on values of the parameter, we prefer to view selection as sorting the total sample according to relative values. Thus, our algorithm will incorporate the idea that those n_c individuals who are highest in terms of some criterion (which may depend on their growth

characteristics) will be assigned to the control group, and the n_p lowest to the program group.

One might argue that we also ought to consider assignment on the basis of sociological or demographic variables. Kenny (1975), using a path analysis model, takes just this approach. While clearly these variables may operate to determine who receives services--for example, income qualifications for compensatory education programs--they are not operant variables in our causal models of human development. Rather, we view these gross observables as reflections of the normally unmeasured individual growth curve parameters. When these growth parameters are available, however, assignment to groups on this basis extracts all of the useful information normally supplied by the gross proxy variables.

The assignment of units to the program and control groups is modelled via the assumption of an assignment variable, A_1. Let us define

$$A_i = \sum_{k=0}^{K} w_k \pi^*_{ki} + w_{K+1} d_i \tag{3.25}$$

where $\pi_{oi} = t_{oi}$

$$\pi^*_{ki} = \frac{\pi_{ki} - \mu_{\pi k}}{\sigma_{\pi k}} = \text{growth parameters in a standardized (unit) metric}$$

$$d_i \sim N(0,1) = \text{random normal deviate} \tag{3.26}$$

and

$$w_{K+1} = 1 - \sum_{k=0}^{K} |w_k| \tag{3.27}$$

(3.28)

The assignment variable, A_i, is a linear combination of two components. The first component is itself a linear combination of the growth parameters. These parameters are expressed now in standardized form, to adjust for potentially different metrics. The second component is a random normal deviate, generated independently for each subject. Any set of weights, [w_k where $(k = 1...K + 1)$], can be specified. This permits total flexibility to model situations ranging from assignment at random ($w_{k+1} = 1$) to assignment based totally on starting point ($w_0 = 1$), or growth rate ($w_\pi = 1$).

Once we have chosen a set of weights, w_k, and sampled for each subject a d_i, we can compute an A_i. If we rank order the subjects on the basis of their values on A_i, we can then assign the bottom $\frac{n_{p_X}}{n_T}$ 100 percent to the program group, and the remainder to the control group. In other than pure random assignment situations, this should be a reasonable replica of real-world selection processes.

Recalling the definition of ϕ_i from the previous section, we can write our assignment algorithm mathematically as

$$A_i \leq A_{(n_p)} \Rightarrow \phi = 1$$

$$A_i > A_{(n_p)} \Rightarrow \phi = 0$$

where $A_{(k)}$ represents the k-th order statistic out of the sample of size n.

4.0 Analysis of the Linear Growth System

4.1 The Linear Model

In this section we analyze a simple form of the general model described above. Our purpose is to provide some general insight into the properties of growth systems and in particular to examine the properties of adjustment techniques.

For the linear system,

$$\frac{\partial G}{\partial t} = \pi_i \tag{4.1}$$

the rate of growth for each subject i is some constant value π_i, but in general the values of the π_i's will vary across subjects to form some distribution. Thus, we have

$$G_i(t) = \pi_i(t-t_{oi}) \text{ where} \tag{4.2}$$

π_i = growth rate for subject i

t_{oi} = date of growth onset

To analyze this model, let us condier the simple case where the treatment effect rate is a constant, ε, for all subjects

$$\frac{\partial I}{\partial t} = \begin{matrix} \varepsilon & t_1 \leq t \leq t_2 \\ o & \text{elsewhere} \end{matrix} \tag{4.3}$$

Thus,

$$\frac{\partial S_i}{\partial t} = \pi_i + \phi_i \varepsilon \text{ and,} \tag{4.4}$$

$$S_i(t) = \pi_i(t-t_{oi}) + \phi_i \varepsilon(t-t_1) \tag{4.5}$$

$$t_1 \leq t \leq t_2$$

At post test time t_2, for subjects in the control group,

$$S_{ic}(t_2) = \pi_i(t_2 - t_{oi}) \text{ and} \tag{4.6}$$

for subjects in the program group,

$$S_{ip}(t_2) = \alpha + \pi_i(t_2 - t_{oi}), \text{ where } \alpha \equiv \varepsilon(t_2 - t_1) \tag{4.7}$$

The full model presented in section 3, can also include a random component $R_i(t)$. Since we are at present only considering the simple classical measurement model (errors have zero mean, constant variance, and are independent of systematic growth) this added complexity would provide little insight. Thus, we will focus in this section only on systematic growth $S_i(t)$.

4.2 Properties of the Linear System

In this section we derive some basic properties of the linear growth model described above. These properties may have applications in the general study of developmental processes. Our main purpose here, however, is to use the results to analyze the linear adjustment strategies described above.

Let

$$E(\pi) = \mu_\pi, \text{ Var } (\pi) = \sigma_\pi^2 \tag{4.8}$$

$$E(t_0) = \mu_{t_0}, \text{ Var } (t_0) = \sigma_\tau^2 \tag{4.9}$$

Then we obtain the following result, which we present as a theorem.

Theorem: For the linear growth system with π and t_0 independent, we have

$$E[G(t)] = \mu_\pi(t - \mu_\tau) \tag{4.10}$$

$$\text{Cov}(G(t_1), G(t_2)) = (t_1 - \mu_\tau)(t_2 - \mu_\tau)\sigma_\pi^2 + \mu_\pi^2\sigma_\tau^2 + \sigma_\pi^2\sigma_\tau^2 \tag{4.11}$$

Proof: Equation (4.10) follows directly from the (4.2) and the independence of π and t_0. From (4.10) and the definition of covariance we have

$$\text{Cov } (G(t_1), G(t_2)) = E\ [\pi^2(t_1-t_0)\ (t_2-t_0) - \mu_\pi^2(t_1-\mu_\tau)\ (t_2-\mu_\tau)]$$

$$= E\ [\pi^2]\ E\ [(t_1-t_0)\ (t_2-t_0)] - \mu_\pi^2(t-\mu_\tau)\ (t_2-\mu_\tau)] \qquad (4.12)$$

Now from the definition of variance, it follows that

$$E(\pi^2) = \sigma_\pi^2 + \mu_\pi^2;\ E(t_0^2) = \sigma_\tau^2 + \mu_\tau^2 \qquad (4.13)$$

Moreover,

$$E\ [(t_1-t_0)\ (t_2-t_0)] = E\ [t_1t_2 - t_2t_0 - t_1t_0 + t_0^2]$$

$$= t_1t_2 - t_2\mu_\tau - t_1\mu_\tau + \sigma_\tau^2 + \mu_\tau^2$$

$$= (t_1-\mu_\tau)\ (t_2-\mu_\tau) + \sigma_\tau^2 \qquad (4.14)$$

Substituting (4.13) and (4.14) into (4.12) and reducing leads to equation (4.11).

From (4.11) it follows that

$$\text{Var } (G(t)) = (t-\mu_\tau)^2\sigma_\pi^2 + \mu_\pi^2\sigma_\tau^2 + \sigma_\pi^2\sigma_\tau^2 \qquad (4.15)$$

Now let ρ_{12} be the correlation between $G(t_1)$ and $G(t_2)$. Then from (4.11) and (4.15) we have

$$\rho_{12} = \frac{(t_1-\mu_\tau)\ (t_2-\mu_\tau)\ \sigma_\pi^2 + \mu_\pi^2\sigma_\tau^2 + \sigma_\pi^2\sigma_\tau^2}{\{[(t_1-\mu_\tau)^2\sigma_\pi^2 + \mu_\pi^2\sigma_\tau^2 + \sigma_\pi^2\sigma_\tau^2]\ [(t_2-\mu_\tau)^2\sigma_\pi^2 + \mu_\pi^2\sigma_\tau^2 + \sigma_\pi^2\sigma_\tau^2]\}^{1/2}} \qquad (4.1$$

The behavior of ρ_{12} as μ_τ, t_1, and t_2 vary is interesting. In particular let us examine a few limiting cases. It is not difficult to show that for fixed $t_1 > \mu_\tau$,

$$\lim_{t_2 \to t_1} \rho_{12} = 1. \qquad (4.17)$$

For fixed $t_2 > \mu_\tau$,

$$\lim_{t_1 \to \mu_\tau} \rho_{12} = \frac{\mu_\pi^2 \sigma_\tau^2 + \sigma_\pi^2 \sigma_\tau^2}{(t_2-\mu_\tau)^2 \sigma_\pi^2 + \mu_\pi^2 \sigma_\tau^2 + \sigma_\pi^2 \sigma_\tau^2}^{\frac{1}{2}} \tag{4.18}$$

For fixed $t_1 > \mu_\tau$,

$$\lim_{t_2 \to \infty} \rho_{12} = \frac{(t_1-\mu_\tau)^2 \sigma_\pi^2}{(t_1-\mu_\tau)^2 \sigma_\pi^2 + \mu_\pi^2 \sigma_\tau^2 + \sigma_\pi^2 \sigma_\tau^2}^{\frac{1}{2}} \tag{4.19}$$

From either (4.18) or (4.19) we also have

$$\lim_{\substack{t_1 \to \mu_\tau \\ t_2 \to \infty}} \rho_{12} = 0 \tag{4.20}$$

In situations of practical interest t_1 will be sufficiently greater than μ_τ to insure that nearly all individuals display positive growth at pre-test time. Thus equation (4.19) is of particular interest. This result can be interpreted as saying that as the interval between pre- and post-tests increases, the correlation decreases to a non-zero minimum value.

4.3 Linear Adjustment Strategies under the Linear Growth Model

So far in this section we have discussed the linear growth system for a single sample of individuals and described some of its important properties. Suppose now that we have two groups, a program group and a control group, as described in section 3. We assume also that the treatment effect can be represented by a constant increment for each individual. As mentioned in section 2, many estimators of the treatment effect which have been suggested are of the form

$$\hat{\alpha} = S_{.p}(2) - S_{.c}(2) - \beta\,(S_{.p}(1) - S_{.c}(1)) \qquad (4.21)$$

where

$S_{.p}(2)$ = post-test mean for program group

$S_{.c}(2)$ = post-test mean for control group

$S_{.p}(1)$ = pre-test mean for program group

$S_{.c}(1)$ = pre-test mean for control group

The value of β may be fixed or a function of the observed values of $S_i(1)$ and $S_i(2)$. For example, gain score analysis corresponds to $\beta = 1$, while for the usual ANCOVA β is the pooled within-group regression of S(2) on S(1).

For a random sample of n individuals, let

$$G_.(t) = \frac{\sum_{i=1}^{n} G_i(t)}{n} \qquad (4.22)$$

Then from (4.10) we have

$$E[G_.(t)] = \mu_\pi(t-\mu_\tau) \qquad (4.23)$$

Let us assume for simplicity that $n_c = n_p = n$. Then we have for the program group

$$E[S_{.p}(1)] = \mu_{\tau p}(t_1-\mu_{\tau p})$$

$$E[S_{.p}(2)] = \alpha + \mu_{\pi p}(t_2-\mu_{\tau p})$$

and for the controls

$$E[S_{.c}(1)] = \mu_{\pi c}(t_1-\mu_{\tau c})$$

$$E[S_{.c}(2)] = \mu_{\pi c}(t_2-\mu_{\tau c})$$

Suppose first that β is fixed. Then

$$E(\hat{\alpha}) = \alpha + \mu_{\pi p}(t_2-\mu_{\tau p}) - \mu_{\pi c}(t_2-\mu_{\tau c}) \qquad (4.24)$$

$$-\beta\mu_{\pi p}(t_1-\mu_{\tau p}) + \beta\mu_{\pi c}(t_1-\mu_{\tau c})$$

Solving for β, we find that $\hat{\alpha}$ will be unbiased if and only if $\beta = \beta^*$, where

$$\beta^* = \frac{\mu_{\pi p}(t_2-\mu_{\tau p}) - \mu_{\pi c}(t_2-\mu_{\tau c})}{\mu_{\pi p}(t_1-\mu_{\tau p}) - \mu_{\pi c}(t_1-\mu_{\tau c})} \tag{4.25}$$

This is a rather remarkable result. It represents the theoretically "correct" adjustment coefficient under our model. Of course, in general, the growth parameters and hence β^* will be unknown.

One special case is of particular interest. Suppose $\mu_{\pi p} = \mu_{\pi c}$. Then the above expression reduces to $\beta^* = 1$. That is, if assignment is effectively on the basis of t_0 but not on π, then gain score analysis provides an unbiased estimate of α.

Suppose now that β is a function of the data. In this case, β is a random variable. Thus $\hat{\alpha}$ is a complex, non-linear function of several correlated random variables. The expectation of such a function is extremely complex. Thus, the exact sampling properties of $\hat{\alpha}$ are more easily investigated via simulation. In the next section we describe a simulation program designed to accomplish this, and we present selected results.

The present analysis, however, can be pushed somewhat further if we turn to asymptotic theory. Since many educational interventions involve rather large samples, these asymptotic results may be of considerable practical importance.

For any random variable, let plim represent the limit in probability as $n \to \infty$. Then from equation (4.21) it follows that

$$\text{plim } \hat{\alpha} = \alpha + \text{plim } S_{.p}(2) - \text{plim } S_{.c}(2) - (\text{plim } \beta)(\text{plim } S_{.p}(1) - \text{plim } S_{.c}(1)) \tag{4.26}$$

Now in general, the probability limits of the sample means will be the population means given above. Thus we have immediately the following

---

Theorem: Under the linear growth model with constant treatment effect α, an estimator of the form

$$\hat{\alpha} = S_{.p}(2) - S_{.c}(2) - \beta(S_{.p}(1) - S_{.c}(1))$$

is a consistent estimate of α if and only if

$$\operatorname*{plim}_{n\to\infty} \beta = \beta^*$$

where

$$\beta^* = \frac{\mu_{\pi p}(t_2-\mu_{\tau p}) - \mu_{\pi c}(t_2-\mu_{\tau c})}{\mu_{\pi p}(t_1-\mu_{\tau p}) - \mu_{\pi c}(t_1-\mu_{\tau c})}$$

Let us now consider the consequences of using some of the β's which have been suggested. One of the simplest is Belson's (1956) suggestion to use the sample regression coefficient based on the control group only. Let β_c represent the population regression coefficient, and $\hat{\beta}_c$ the sample estimate. Then since sample moments approach the corresponding population values, we have

$$\operatorname{plim} \hat{\beta}_c = \beta_c.$$

Moreover, from equations (4.11) and (4.12) we have

$$\beta_c = \frac{(t_1-\mu_{\tau c})(t_2-\mu_{\tau c})\sigma^2_{\pi c} + \mu^2_{\pi c}\sigma^2_{\tau c} + \sigma^2_{\pi c}\sigma^2_{\tau c}}{(t_1-\mu_{\tau c})^2\sigma^2_{\pi c} + \mu^2_{\pi c}\sigma^2_{\tau c} + \sigma^2_{\pi c}\sigma^2_{\tau c}}$$

In general, $\beta_c \neq \beta^*$, so that the Belson method is not under our model a consistent estimate. Similarly, it can be shown that

none of the other suggested β's are in general consistent. Of course, this result should not be surprising, since the linear adjustment estimators were not derived on the basis of an explicit model of growth such as the one presented here. On the other hand, it is somewhat disconcerting that under a reasonable, though overly simplified, model of growth, the estimators would be misleading even in very large samples.

Another aspect of the linear adjustment problem pertains to the fact that the sample regression coefficients in the two groups will in general converge to different values. Suppose β_p represents the population regression coefficient for the program group and $\hat{\beta}_p$ the sample coefficient. Then

$$\text{plim } \hat{\beta}_p = \beta_p$$

where

$$\beta_p = \frac{(t_1-\mu_{\tau p})\ (t_2-\mu_{\tau p})\sigma^2_{\pi p} + \mu^2_{\pi p}\sigma^2_{\tau p} + \sigma^2_{\pi p}\sigma^2_{\tau p}}{(t_1-\mu_{\tau p})^2\sigma^2_{\pi p} + \mu^2_{\pi p}\sigma^2_{\tau p} + \sigma^2_{\pi p}\sigma^2_{\tau p}} \tag{4.28}$$

Thus in general

$$\beta_p \neq \beta_c \tag{4.29}$$

This suggests that if our model is approximately correct in a real situation, the data analyst is likely to encounter "heterogeneity of regression". Thus heterogeneity of regression may turn out to have the character not of a minor impediment which rarely occurs in practice, but rather of a problem which is endemic to data generated in growth situations.

There is one special, though rather trivial, case where $\beta_c = \beta_p = \beta^*$. This occurs when $\sigma^2_{\tau c} = \sigma^2_{\tau p} = 0$ and $\mu_{\tau p} = \mu_{\tau c} = t_0$. That is, all individuals in both groups "fan out" from the same starting point t_0. In that case $\beta_c = \beta_p = \beta^* = \frac{t_2 - t_0}{t_1 - t_0}$. Moreover, it can be shown that the coefficients specified by standardized gains and ANCOVA, as well as the Belson approach all reduce to this value. Thus, linear adjustments can be expected to perform well in situations approximated by this model.

On the other hand, suppose $\mu_{\tau p} \neq \mu_{\tau c}$ but that $\mu_{\pi p} = \mu_{\pi c}$. We showed above that gain scores ($\beta^* = 1$) would be appropriate in this situation. From (4.27) and (4.28) however, we have $\beta_c \neq \beta_p$, and both β_c and β_p are greater than 1. Thus we see that the argument which favors the use of ANCOVA over gain scores on the grounds that the former is a special case of the latter is faulty. It turns out ironically that for models where gain scores are theoretically correct, ANCOVA as an estimation technique does poorly. Thus the data analyst cannot rely on ANCOVA to select β automatically if it happens to be near 1.

5.0 Computer Simulation

5.1 The Simulation Program

In the previous section we developed some analytic results for the linear case. The component for the assignment model, however, is particularly difficult to manipulate analytically. Moreover, conclusions cannot easily be drawn for the more complicated growth and treatment effect models or for the non-linear adjustment strategies such as standardization. Thus, we resort to simulation.

We have developed a general FORTRAN IV computer program to generate and analyze data under a broad range of conditions as described in sections 2 and 3. Our main purpose is to assess the ability of various adjustment strategies to remove bias in quasi-experimental studies involving growth systems. For each set of conditions (a growth system, a treatment effect model, and an assignment procedure), we define the theoretical treatment effect by

$$\alpha = E_p[S(t_2) - G(t_2)] \tag{5.1}$$

where the expectation is over the distribution of individuals assigned to the program group. For each analysis strategy there is an estimator $\hat{\alpha}$. Thus, we can define the bias of any analysis strategy as

$$\text{BIAS}(\hat{\alpha}) = E(\hat{\alpha}) - \alpha \tag{5.2}$$

Two features built into the simulation program are worthy of comment.[1] First, we place certain constraints on our simulations in order to generate "reasonable" data. The basic statistics for

each generated data set are examined according to a set of reasonableness criteria.[2] If the data set fails to pass any one of the criteria, it is rejected from the simulation run. If a particular simulation run (i.e., a growth system, an assignment procedure, and a treatment effect model) has more than 5% rejected data sets, the run is terminated. In this manner, we seek to insure that conditions utilized in the simulation may be a reasonable reflection of data sets seen in common practice.

Second, a flexible simulation termination procedure is built into the program. A maximum number of runs is specified, but the program can be terminated if the estimate of bias achieves sufficient precision. We specified the maximum number of runs to be 350.

We decided to let the simulation terminate when a given tolerance was reached, i.e.,

$$\Pr\{|\bar{\alpha} - E(\alpha)| \geq h\} < .01 \qquad (5.3)$$

From the normal approximation, we can determine the maximum number of runs (NRUNS) necessary for any strategy j:

$$NRUNS_j = \left(\frac{2.58}{h}\right)^2 S_j^2 \qquad j = 1, 2, \ldots 16 \qquad (5.4)$$

The tolerance must be specified external to the program. We selected

$$h = .3 \qquad (5.5)$$

for all runs. The program estimates S_j^2 for each strategy after 100 runs. The value of $NRUNS_j$ is calculated for each strategy.

The number of runs to termination is then calculated as:

$$N = \min [350, \max_{j} (NRUNS_j)] \qquad (5.6)$$

Finally, we let $n_c = n_p = 60$ in all cases. We have also run a smaller sample case $n_c = n_p = 30$ but the results for bias turn out, not surprisingly, to be independent of sample size for most strategies. Of course, the precision will be related to sample size, but since the focus of this paper is bias, we have omitted the small sample results.

5.2 Growth Models Considered

So far we have carried out runs only for the linear growth model described in Section 4. However, we have examined the consequences of eight different sets of growth conditions (see Table 3) crossed with six assignment models (see Table 4) and three treatment effect models.

One treatment effect model is the constant effect discussed in section 4 (see equation 4.5). In addition, we have considered two interactive models. The first represents a situation where the treatment effect is largest for children with the lowest growth status. In terms of equation (3.12), we can represent this model as [3]

$$\frac{\partial I_i}{\partial t} = \varepsilon_i(100 - G_i(t)) = \varepsilon_i(100 - \pi_i(t-t_{oi})) \qquad (5.7)$$

Thus for linear growth

$$\frac{\partial S_i}{\partial t} = \pi_i + \phi_i \varepsilon_i(100 - \pi_i(t-t_{oi})) \qquad (5.8)$$

Integration yields after reduction

$$S_i(t) = \pi_i(t-t_{oi}) + \phi_i \varepsilon_i(t-t_1) [100-\pi_i \frac{(t + t_1)}{2} + t_{oi}] \quad (5.9)$$

Thus at post-test time t we have a variable treatment effect:

$$\alpha_i = \varepsilon_i(t_2-t_1) [100-\pi_i \frac{(t_1+t_2)}{2} + t_{oi}] \quad (5.10)$$

A third model represents the situation where the treatment effect is largest for children with the highest status. This model can be represented as

$$\frac{\partial I_i}{\partial t} = \varepsilon_i G_i(t) \quad (5.11)$$

which leads to

$$S_i(t) = \pi_i(t-t_{oi}) + \phi_i \pi_i(1 + \varepsilon_i) [\frac{t^2-t_1^2}{2} - t_{oi}] \quad (5.12)$$

and

$$\alpha_i = \pi_i(1 + \varepsilon_i) [\frac{t_2^2 - t_1^2}{2} - t_{oi}] \quad (5.13)$$

In the theoretical development in section 4, we did not include random growth in the model. In our simulation models we have assumed in all cases that there exists a random component $R_i(t)$ such that for any t, R is independent of S and

$$R_i(t) \sim N(0, 5). \quad (5.14)$$

5.3 Simulation Results

In tables 5 through 16 we present the biases for each of the 16 analysis strategies defined in Table 1. The first row of each table represents the bias which would result with no statistical adjustment, i.e., if the simple difference of post-test means were used to estimate the treatment effect. Each

of the remaining fifteen rows corresponds to the strategy with the same number in Table 1. A negative sign indicates an underestimate of the true effect.

Tables 5 through 10 present results for the linear case for each of the six assignment procedures considered. Tables 11 through 16 present results for the "compensatory effect" model in which lower starting individuals get the largest boost. The pattern of results for the "rich get richer" model is similar, and we have for the sake of brevity decided to omit these tables.

Several interesting patterns can be observed in the simulation results:

(1) As predicted in section 4, non-random assignment on the basis of t_o but not on π, results in substantial bias for all techniques except gain scores (strategy 2).

(2) The value-added technique (strategy 10) performs best in general over the range of conditions considered here. The Belson ANCOVA with reliability correction based on both age and pre-test (strategy 15) also does well. The need to know the pre-test reliability, however, is a limitation of this approach.

(3) When the treatment effect interacts with individual characteristics, strategies based primarily on the pre-post relationship in the control group only, (strategies 7, 8, 14, 15) perform consistently better than the corresponding strategies based on relationships in both groups (strategies 5, 6, 12, 13). This makes sense intuitively, since the interactive treatment will alter the pre-post relationship in the program group but not the control group.

(4) When assignment of subjects to groups is based on either (but not both) of the growth parameters (assignment sets 1 and 2) or on both operating to create bias in the same direction (assignment set 6), all adjustment strategies reduce the initial bias to some extent.

(5) When the assignment is based on both π and t_o in such a way that they tend to create biases in opposite directions (assignment sets 3, 4, 5), no adjustment strategy does well. In fact, almost all strategies considered do worse across all the growth conditions.

6.0 Summary and Conclusions

As mentioned in the introduction, social research and evaluation methodologists disagree on whether statistical adjustments are adequate to remove bias in quasi-experimental situations. We have suggested that such a debate can move towards resolution only through the consideration of precisely stated models which realistically represent the way observational data are generated. This paper represents a step in this direction.

We have developed a model to represent somewhat realistically the way quasi-experimental data are generated when dynamic, developmental processes are being studied. This model includes explicitly three important components of the quasi-experimental situation: the nature of individual growth, the nature of the treatment effect, and the process by which individuals are assigned to groups.

Our goal in this work has been to understand the consequences, under our general model, of employing the various adjustment strategies which have been suggested. In section 4, we assumed a particularly simple form of the model: linear growth and constant treatment effect. These assumptions allowed us to analyze the properties of the model in some detail; and we were able to obtain several suggestive results.

For one thing we found that under our model, none of the usual techniques yields a consistent estimator of the true treatment effect. In practical terms, this means that if the linear growth model with constant treatment effect is approximately correct, then even in very large samples the statistical adjustment strategies will be somewhat misleading.

Further, we found that except in a very special case (when all individuals in a group "fan out" from a common starting point), the regression coefficient of post-test on pre-test will differ in the two treatment groups. Thus "heterogeneity of regression" is not accidental, but is built into this model. In practical terms this means that if the linear growth model with constant treatment effect is approximately correct, then heterogeneity of regression is likely to be observed in the data. Thus linear adjustments will not be deemed appropriate.

Finally, we showed that although there are models where gain score analysis is appropriate, and models where traditional analysis of covariance is appropriate, there is no model under which both will work well. In practical terms this means that without some judgment by the data analyst on the nature of the underlying growth model, the justification for either gain scores or ANCOVA is weak.

These results are suggestions, not conclusions, since they are based on a particular model whose realism is somewhat questionable. On the other hand, there is no obvious reason to believe that more complex models will make the case for statistical adjustments any stronger. To explore the properties

of some more complex models we developed the computer simulation program described in section 5. Simulation was used to explore the linear model with constant treatment described above, and two models involving non-constant, interactive treatment effects described in section 5. Non-linear growth models are presently under investigation but the results are not yet ready for presentation.

Several general findings emerged from the simulation results, and were presented in section 5. Summarizing these results, we can say that no technique based on either the pre-test alone or the pre-test and age, consistently removes most of the bias. Moreover, there are assignment processes for which all of the strategies considered do consistently worse than the unadjusted post-test mean comparison. In practical terms, adjustment techniques may make comparisons more biased or less biased, depending on the nature of growth, treatment effect, and assignment to groups.

In summary, it seems fair to say that statistical adjustment techniques are not adequate to eliminate bias resulting from non-random assignment in comparative studies involving growth systems. If one accepts this general conclusion, there are two possible implications. One is that randomization is absolutely essential in social and educational evaluations. In our view, this conclusion is unacceptable. As a practical matter, even under the most optimistic assumptions, randomization could be employed only in a small fraction of evaluations.

An alternative conclusion is that the classical pre-test and post-test, treatment versus control design should not be used in comparative studies involving growth. There is simply not enough information generated by this design to assess the effect of an intervention, without supplementary information on the form of growth, the nature of the treatment effect, and the assignment of subjects to groups.

In terms of future research, this perspective points in several directions. First, more realistic models of educational and social growth processes must be developed. Such models must include parameters which are meaningful in terms of individual growth. The effect of an intervention could be described in terms of changes in these individual growth-relevant parameters.

Hand in hand with the development of better models must go development of experimental designs and statistical methods which will allow the estimation of model parameters. Such designs will undoubtedly involve longitudinal data on individuals collected at more than two time points.

The static production function model of educational and social processes seems to us inadequate. Linear additive models and the corresponding experimental and quasi-experimental designs which allow estimation of the paramtersof these models are therefore not very useful in studying dynamic social and educational situations. To randomize or not to randomize is not the most important question. Of primary importance is the appropriateness of the model assumed.

Footnotes

[1] For a comprehensive description of the entire program see Appendix B.

[2] Appendix C contains the FORTRAN program used to generate and analyze the simulation data. The actual reasonableness criteria utilized can be found in the NOKNOK subroutine.

[3] In the simulations the data are scaled to be between 0 and 100 boundaries. We chose 100 rather than 1 so that the "scores" generated would be more realistic. Also, growth conditions were selected which maintain the integrity of these boundaries. That is, we wished to avoid floor and ceiling effects.

Bibliography

1. Bock, D.R., Wainer, H., Peterson, A., Thissen, D., Murray, J. and Roche, Alex. A Parameterization for Individual Human Growth Curves, Human Biology, 1973. Vol. 45.

2. Bryk, A.S. and Weisberg, H.I. A New Approach to Analyzing Quasi-Experimental Data: Value-Added Analysis. Proceedings of the American Statistical Association, Social Statistics Section, 1974.

3. Bryk, A.S. and Weisberg, H.I. Value-Added Analysis: A Dynamic Approach to the Analysis of Quasi-Experiments. Journal of Educational Statistics, 1976 (forthcoming).

4. Campbell, D.T. and Erlebacher, A. How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful. In J. Hellmuth (Ed.), Compensatory Education: A National Debate. Vol. III. Disadvantaged Child. New York: Brunner-Mazel, 1970.

5. Cochran, W.G. The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies. Biometrics, 24, 1968.

6. Cochran, W.G. and Rubin, D.B. Controlling Bias to Observational Studies: A Review. Sankhya, The Indian Journal of Statistics, Series A. Volume 35, Part 4, 1973.

7. Cronbach, L.J. and Furby, L. How We Should Measure "Change"--Or Should We? Psychological Bulletin. 1970, 74.

8. Gilbert, J.P., Light, R.J. and Mosteller, F. Analyzing Social Programs: An Empirical Base for Social Policy. In Arthur Lunsdale and Carl Bennett (Eds.). Experiment and Evaluation. Academic Press, 1975.

9. Kenny, D. A Quasi-Experimental Approach to Assessing Treatment Effects in the Nonequivalent Control Group Design. Psychological Bulletin, Vol. 76, 1975.

10. Lord, F.J. A Paradox in the Interpretation of Group Comparisons. Psychological Bulletin, 1967.

11. Riecken, H.W. and Boruch, R.F. (Eds.) Social Experimentation: A Method for Planning and Evaluating Social Intervention. New York: Academic Press, 1975.

12. Rubin, D.B. The Use of Matched Sampling and Regression Adjustment to Remove Bias in Observational Studies. Biometrics, Vol. 29, No. 1, 1973.

13. Smith, P. Estimating Linear Trends in Contaminated Quasi-Experiments of Pre-test-Posttest Design: or: Gambling on Linear Models, 1972. (Unpublished).

14. Tucker, L.R., Damarin, F. and Messick, S. A Base-free Measure of Change. Psychometrika, 1966, 31.

15. Werts, C.E. and Linn, R.L. A General Linear Model for Studying Growth. Psychological Bulletin, 1970, 73.

16. Wiley, D.E. Aproximations to ceteris paribus: Data Adjustment in Educational Research. Paper presented at International Research Seminar on Learning and the Educational Process, Munich, Germany, 1971.

Appendix A

Tables 1 through 16

Table 1

The Lineup

(Analysis Strategies)

1)	Post-test only	$\bar{Y}_{2p} - \bar{Y}_{2c}$
2)	Raw gains	$\bar{Y}_{2p} - \bar{Y}_{2c} - (\bar{Y}_{1p} - \bar{Y}_{1c})$
3)	Standardized gains	$\bar{Y}_{2p} - \bar{Y}_{2c} - \frac{s_2}{s_1} (\bar{Y}_{1p} - \bar{Y}_{1c})$
4)	Residual analysis (pre-test only)	$\bar{Y}_{2p} - \bar{Y}_{2c} - \beta_1 (\bar{Y}_{1p} - \bar{Y}_{1c})$
5)	ANCOVA (pre-test only)	$\bar{Y}_{2p} - \bar{Y}_{2c} - \beta_2 (\bar{Y}_{1p} - \bar{Y}_{1c})$
6)	ANCOVA with correction (pre-test only)	$\bar{Y}_{2p} - \bar{Y}_{2c} - \frac{\beta_2}{\rho} (\bar{Y}_{1p} - \bar{Y}_{1c})$
7)	Belson ANCOVA (pre-test only)	$\bar{Y}_{2p} - \bar{Y}_{2c} - \beta_3 (\bar{Y}_{1p} - \bar{Y}_{1c})$
8)	Belson ANCOVA with ρ correction (pre-test only)	$\bar{Y}_{2p} - \bar{Y}_{2c} - \frac{\beta_3}{\rho} (\bar{Y}_{1p} - \bar{Y}_{1c})$
9)	Subclassification with 6 categories (pre-test only)	
10)	Value-Added	$\bar{Y}_{2p} - \bar{Y}_{1p} - \eta(\bar{a}_{2p} - \bar{a}_{1p})$
11)	Residual analysis	$\bar{Y}_{2p} - \bar{Y}_{2c} - \beta_4 (\bar{Y}_{1p} - \bar{Y}_{1c}) - \gamma_2 (\bar{a}_{1p} - \bar{a}_{1c})$
12)	ANCOVA (pre-test and age)	$\bar{Y}_{2p} - \bar{Y}_{2c} - \beta_5 (\bar{Y}_{1p} - \bar{Y}_{1c}) - \gamma_3 (\bar{a}_{1p} - \bar{a}_{1c})$
13)	ANCOVA with correction (pre-test and age)	$\bar{Y}_{2p} - \bar{Y}_{2c} - \beta_6 (\bar{Y}_{1p} - \bar{Y}_{1c}) - \gamma_4 (\bar{a}_{1p} - \bar{a}_{1c})$
14)	Belson ANCOVA (pre-test and age)	$\bar{Y}_{2p} - \bar{Y}_{2c} - \beta_7 (\bar{Y}_{1p} - \bar{Y}_{1c}) - \gamma_5 (\bar{a}_{1p} - \bar{a}_{1c})$
15)	Belson ANCOVA with ρ correction (pre-test and age)	$\bar{Y}_{2p} - \bar{Y}_{2c} - \beta_8 (\bar{Y}_{1p} - \bar{Y}_{1c}) - \gamma_6 (\bar{a}_{1p} - \bar{a}_{1c})$
16)	Two-way Subclassification (pre-test by age)	

Table 2

Glossary of Terms in Table 1

1) ρ Reliability of pre-test

2) s_1 Pooled within group estimated standard deviation of pre

3) s_2 Pooled within group estimated standard deviation of post

4) β_1 Regression coefficient of post on pre for entire sample

5) β_2 Pooled within group regression coefficient of post on pre

6) β_3 Regression coefficient of post on pre based on control group only

7) β_4 Coefficient of post on pre in multiple regression with pre and age, based on entire sample

8) β_5 Coefficient of post on pre in multiple regression with pre and age, based on pooled within group covariance matrix

9) β_6 Same as β_5 except that covariance matrix is corrected for attenuation

10) β_7 Coefficient of post on pre in multiple regression with pre and age, based on control group only

11) β_8 Same as β_7 except that covariance matrix is corrected for attenuation

12) γ_1 Regression coefficient of pre on age based on program group only.

13) γ_2 Coefficient of post on age in multiple regression with pre and age, based on entire sample

14) γ_3 Coefficient of post on age in multiple regression with pre and age, based on pooled within group covariance matrix.

15) γ_4 Same as γ_3 except that covariance matrix is corrected for attenuation

16) γ_5 Coefficient of post on age in multiple regression with pre and age, based on control group only

17) γ_6 Same as γ_5 except that covariance matrix is corrected for attenuation

Table 3

Growth Parameter Conditions for the Linear System

Growth Conditions	1	2	3	4	5	6	7	8
μ_d	9	9	9	9	9	9	9	9
σ_d^2	.6	.3	.6	.3	.6	.3	.6	.3
μ_π	2.5	2.5	2.5	2.5	5.0	5.0	5.0	5.0
σ_π^2	.4	.4	.8	.8	.4	.4	.8	.8

Distributions of d and π are assumed to be normal.

Distribution of b is uniform (0,6) for all cases.

Table 4

Non Random Assignment Conditions

Considered in this Simulation

Weights			
$w_{t_{oi}}$	w_{π_i}	w_{d_i}	Process Description of Individuals Assigned to the Control Group
-.2	0	.8	early starters
0	.2	.8	fast growers
.4	.4	.2	late starters, but fast growers
.2	.4	.4	late starters, but fast growers
.4	.2	.4	late starters, but fast growers
-.2	.2	.6	early starters and fast growers

Table 5

Bias Associated with Analysis Strategies

Model

Growth model: linear

Treatment effect: constant

Assignment: $w_{t_o} = -.20$ $w_{\pi} = 0.0$ $w_d = .80$

Analysis	Growth Conditions							
	1	2	3	4	5	6	7	8
1	-1.52	-1.39	-1.33	-1.32	-2.78	-2.88	-3.31	-3.03
2	- .02	- .02	- .01	.08	.00	- .09	- .32	- .16
3	.88	.84	1.16	1.33	.64	.59	.79	.98
4	.04	.04	.42	.54	-.12	-.19	-.16	.00
5	.35	.33	.67	.80	.22	.14	.18	.35
6	.68	.66	.94	1.12	.37	.30	.36	.53
7	.41	.39	.69	.82	.26	.19	.29	.46
8	.75	.73	.97	1.14	.42	.35	.48	.64
9	.51	.57	1.00	1.10	.21	.29	.18	.39
10	.27	.14	.19	.25	.12	-.23	-.08	-.13
11	-.30	-.41	-.24	-.24	-.30	-.48	-.56	.47
12	.16	.06	.19	.22	.14	-.02	-.08	.00
13	.39	.13	.30	.30	.28	.06	.06	.11
14	.20	.08	.17	.20	.06	-.01	-.02	.05
15	.42	.19	.29	.29	.30	.08	.11	.16
16	.14	.17	.35	.39	-.15	-.21	-.31	-.18

Table 6

Bias Associated with Analysis Strategies

Model

Growth model: linear

Treatment effect: constant

Assignment: $w_{t_o} = 0.0$ $w_{\pi} = .20$ $w_{d} = .80$

Analysis	Growth Conditions							
	1	2	3	4	5	6	7	8
1	-3.09	-2.88	-4.05	-4.09	-2.87	-2.90	-4.14	-4.29
2	-2.02	-1.92	-2.72	-2.76	-1.90	-1.99	-2.91	-2.90
3	-1.42	-1.36	-1.61	-1.61	-1.69	-1.78	-2.46	-2.36
4	-1.95	-1.85	-2.26	-2.30	-1.94	-2.03	-2.82	-2.77
5	-1.78	-1.71	-2.09	-2.11	-1.84	-1.92	-2.71	-2.65
6	-1.56	-1.49	-1.82	-1.82	-1.78	-1.87	-2.64	-2.57
7	-1.79	-1.72	-2.13	-2.15	-1.83	-1.90	-2.70	-2.65
8	-1.56	-1.49	-1.87	-1.86	-1.77	-1.85	-2.62	-2.56
9	-1.56	-1.49	-2.03	-1.92	-1.68	-1.74	-2.56	-2.53
10	- .99	- .79	-1.31	-1.37	- .83	-1.07	-1.36	-1.40
11	-1.88	-1.77	-1.97	-1.94	-1.93	-1.93	-2.47	-2.30
12	-1.55	-1.46	-1.62	-1.57	-1.68	-1.62	-2.18	-1.98
13	- .80	- .61	- .93	- .73	-1.41	-1.17	-1.78	-1.45
14	-1.57	-1.45	-1.66	-1.58	-1.68	-1.63	-2.20	-2.00
15	.70	- .42	- .89	- .60	-1.40	-1.13	-1.75	-1.41
16	-1.97	-1.87	-2.29	-2.16	-2.03	-2.26	-2.96	-2.88

Table 7

Bias Associated with Analysis Strategies

Model

Growth model: linear

Treatment effect: constant

Assignment: $w_{t_o} = .40$ $w_{\pi} = .40$ $w_d = .20$

	Growth Conditions							
Analysis	1	2	3	4	5[1]	6[2]	7	8
1	-4.51	-4.36	-8.17	-7.83			-3.96	-3.65
2	-5.87	-5.86	-8.35	-8.11			-8.40	-8.23
3	-6.63	-6.75	-8.49	-8.34			-10.17	-10.20
4	-6.41	-6.49	-8.47	-8.30			-9.73	-9.72
5	-6.29	-6.35	-8.44	-8.36			-9.61	-9.58
6	-6.60	-6.73	-8.47	-8.33			-9.90	-9.89
7	-6.21	-6.24	-8.24	-8.23			-9.42	-9.35
8	-8.51	-6.60	-8.45	-8.29			-9.71	-9.65
9	-5.89	-5.95	-8.28	-8.11			-9.21	-9.21
10	-2.86	-2.81	-3.89	-3.98			-4.00	-4.23
11	-7.99	-8.47	-9.09	-9.20			-9.62	-9.56
12	-6.50	-6.64	-7.76	-7.44			-8.85	-8.39
13	-4.53	-2.46	-5.47	-3.99			-7.94	-6.78
14	-6.78	-7.15	-7.93	-7.54			-8.94	-8.54
15	-3.01	-2.56	-4.02	.05			-7.58	-5.88
16	-7.88	-8.40	-9.94	-10.36			-16.15	-10.49

[1] 29 runs.

[2] 39 runs.

Table 8

Bias Associated with Analysis Strategies

Model

Growth model: linear

Treatment effect: constant

Assignment: $w_{t_o} = .20$ $w_{\pi} = .40$ $w_d = .40$

	Growth Conditions							
Analysis	1	2	3	4	5	6	7	8
1	-6.23	-6.33	-9.77	-9.56	-4.34	-4.54	-7.76	-7.63
2	-5.51	-5.55	-7.83	-5.52	-5.48	-7.78	-7.77	-7.77
3	-5.16	-5.14	-6.43	-6.26	-5.77	-5.69	-7.80	-7.83
4	-5.44	-5.47	-7.17	-7.00	-5.70	-5.65	-7.83	-7.83
5	-5.37	-5.38	-7.02	-6.85	-5.62	-5.58	-7.79	-7.80
6	-5.21	-5.20	-6.65	-6.45	-5.69	-5.63	-7.79	-7.81
7	-5.39	-5.41	-7.18	-7.02	-5.61	-5.55	-7.77	-7.77
8	-5.24	-5.23	-6.83	-6.64	-5.67	-5.61	-7.78	-7.78
9	-5.19	-5.20	-7.24	-7.04	-5.41	-5.45	-7.70	-7.75
10	-2.54	-2.66	-3.64	-3.79	-2.52	-2.75	-4.02	-3.80
11	-6.00	-6.11	-6.88	-6.70	-5.92	-6.00	-7.33	-7.19
12	-5.03	-4.93	-5.76	-5.34	-5.33	-5.16	-6.68	-6.20
13	-3.22	-2.29	-3.79	-2.71	-4.71	-4.09	-5.68	-4.68
14	-5.15	-5.05	-5.79	-5.26	-5.39	-5.22	-6.80	-6.24
15	-2.38	- .42	-2.80	- .68	-4.60	-3.79	-5.52	-4.00
16	-6.29	-6.40	7.97	-7.91	-6.26	-6.51	-8.52	-8.62

Table 9

Bias Associated with Analysis Strategies

Model

Growth model: linear

Treatment effect: constant

Assignment: $w_{t_o} = .40$ $w_{\pi} = .20$ $w_d = .40$

	Growth Conditions							
Analysis	1	2	3	4	5[1]	6[2]	7[3]	8[4]
1	-.02	-.11	-1.66	-1.64				
2	-2.80	-2.90	-3.93	-4.04				
3	-4.56	-4.73	-5.96	-6.26				
4	-4.19	-4.34	-5.51	-5.76				
5	-3.66	-3.80	-5.24	-5.47				
6	-4.30	-4.50	-5.73	-6.04				
7	-3.45	-3.59	-5.03	-5.22				
8	-4.05	-4.25	-5.49	-5.75				
9	-2.99	-3.21	-4.58	-4.83				
10	-1.23	-1.32	-1.77	-1.99				
11	-4.86	-5.19	-5.37	-5.65				
12	-3.04	-2.98	-3.49	-3.43				
13	-2.33	-1.75	-2.61	-2.13				
14	-3.11	-3.08	-3.38	-3.26				
15	-1.91	- .82	-1.94	- .96				
16	-3.76	-4.29	-4.89	-5.32				

[1] 12 runs.

[2] 11 runs

[3] 54 runs.

[4] 20 runs.

Table 10

Bias Associated with Analysis Strategies

Model

Growth model: linear

Treatment effect: constant

Assingment: $w_{t_o} = -.20$ $w_{\pi} = .20$ $w_d = .60$

	Growth Conditions							
Analysis	1	2[1]	3	4	5	6	7	8
1	-5.41		-6.84	-6.80	-7.17	-7.07	-8.53	-8.52
2	-2.44		-3.41	-3.48	-2.46	-2.48	-3.51	-3.42
3	- .76		- .52	- .57	-1.49	-1.43	-1.71	-1.48
4	-2.57		-2.63	-2.68	-2.82	-2.82	-3.38	-3.30
5	-1.86		-1.87	-1.93	-2.22	-2.20	-2.78	-2.65
6	-1.23		-1.19	-1.20	-1.95	-1.95	-2.48	-2.34
7	-1.83		-1.94	-2.00	-2.18	-2.14	-2.73	-2.59
8	-1.20		-1.27	-1.29	-1.92	-1.88	-2.42	-2.27
9	-2.03		-2.34	-2.35	-2.46	-2.47	-3.12	-3.13
10	-1.10		-1.53	-1.56	-1.19	-1.24	-1.69	-1.82
11	-2.59		-2.71	-2.76	-2.86	-2.75	-3.12	-3.02
12	-1.78		-1.86	-1.94	-2.08	-2.01	-2.38	-2.21
13	- .72		- .91	- .86	-1.56	-1.38	-1.73	-1.45
14	-1.80		-2.02	-2.10	-2.06	-1.98	-2.41	-2.26
15	- .74		-1.07	-1.03	-1.55	-1.36	-1.76	-1.52
16	-2.64		-2.93	-3.09	-3.21	-3.23	-3.88	-3.78

[1] 100 runs.

Table 11

Bias Associated with Analysis Strategies

Model

Growth model:

Treatment effect: interactive (largest effect for lowest status)

Assignment: $w_{t_o} = -.20$ $w_{\pi} = 0.0$ $w_d = .80$

	Distributions for Growth Parameters							
Analysis	1	2	3	4	5	6	7	8
1	-1.42	-1.51	-1.50	-1.42	-2.85	-2.96	-2.93	-3.17
2	.05	- .10	- .12	- .03	- .05	- .11	- .02	- .08
3	1.05	.91	1.08	1.23	.45	.42	.91	.97
4	- .24	- .36	- .01	.12	- .56	- .64	- .26	- .33
5	.25	.09	.37	.51	- .15	- .22	.16	.14
6	.54	.39	.63	.80	- .00	- .08	.32	.31
7	.47	.30	.58	.71	.22	.17	.59	.58
8	.80	.64	.86	1.03	.38	.33	.78	.78
9	.51	.39	.81	.96	.24	.21	.51	.53
10	.19	.18	.38	.02	.03	- .11	.29	- .20
11	- .68	- .95	- .71	- .80	- .78	- .89	- .66	- .93
12	.05	- .20	- .05	- .07	- .23	- .33	- .11	- .29
13	.23	- .10	.07	- .03	- .10	- .22	.02	- .21
14	.25	- .02	.11	.08	.12	.02	.27	.06
15	.44	.09	.24	.14	.27	.13	.40	.16
16	.04	- .18	.02	.11	- .22	- .32	- .14	- .07

Table 12

Bias Associated with Analysis Strategies

Model

Growth model: linear

Treatment effect: interactive (largest effect for lowest status)

Assignment: $w_{t_o} = 0$ $w_{\pi} = .20$ $w_d = .80$

	Distributions for Growth Parameters							
Analysis	1	2	3	4	5	6	7	8
1	-2.82	-3.02	-4.21	-4.04	-2.85	-2.85	-4.11	-4.06
2	-1.85	-1.93	-2.81	-2.66	-1.93	-1.89	-2.83	-2.70
3	-1.23	-1.20	-1.61	-1.44	-1.77	-1.72	-2.44	-2.26
4	-2.01	-2.14	-2.65	-2.50	-2.11	-2.08	-2.91	-2.78
5	-1.75	-1.82	- .232	-2.16	-1.97	-1.93	-2.77	-2.62
6	-1.56	-1.60	-2.06	-1.87	-1.92	-1.88	-2.70	-2.54
7	-1.63	-1.70	-2.19	-2.01	-1.85	-1.81	-2.60	-2.44
8	-1.42	-1.44	-1.91	-1.71	-1.80	-1.76	-2.52	-2.35
9	-1.43	-1.49	-2.04	-1.92	-1.67	-1.67	-2.51	-2.34
10	- .72	- .90	-1.27	-1.12	- .58	- .88	-1.40	-1.37
11	-2.04	-2.18	-2.46	-2.28	-2.11	-2.01	-2.62	-2.36
12	-1.52	-1.55	-1.82	-1.61	-1.78	-1.64	-2.24	-1.93
13	- .89	- .68	-1.15	- .82	-1.51	-1.25	-1.86	-1.42
14	-1.40	-1.45	-1.63	-1.42	-1.68	-1.54	-2.08	-1.76
15	- .58	- .32	- .81	- .43	-1.37	-1.08	-1.63	-1.15
16	-2.03	-2.20	-2.65	-2.36	-2.23	-2.15	-2.94	-2.81

Table 13

Bias Associated with Analysis Strategies

Model

Growth model: linear

Treatment effect: interactive (largest effect for lowest status)

Assignment: $w_{t_o} = .40$ $w_{\pi} = .40$ $w = .20$

	Distributions for Growth Parameters							
Analysis	1	2	3	4	5[1]	6[2]	7	8
1	-4.55	-4.37	-8.11	-7.91			-3.81	-3.47
2	-5.86	-5.79	-8.30	-8.17			-8.32	-8.15
3	-6.68	-6.72	-8.44	-8.38			-9.76	-9.83
4	-6.35	-6.37	-8.46	-8.36			-9.31	-9.32
5	-6.08	-6.06	-8.35	-8.26			-9.00	-8.99
6	-6.35	-6.38	-8.39	-8.32			-9.28	-9.28
7	-6.17	-6.14	-8.35	-8.27			-9.35	-9.31
8	-6.45	-6.47	-8.38	-8.33			-9.64	-9.62
9	-5.88	-5.88	-8.19	-8.17			-9.12	-9.24
10	-2.64	-2.78	-3.78	-4.11			-4.22	-3.97
11	-9.75	-10.61	-11.57	-12.15			-9.72	-9.92
12	-6.21	-6.30	-7.75	-7.58			-8.16	-7.77
13	-4.42	-3.50	-5.65	-4.49			-7.29	-6.30
14	-6.67	-6.93	-7.88	-7.63			-8.81	-8.53
15	-1.08	.37	-3.80	.44			-7.42	-5.92
16	-8.98	-9.51	-11.52	-12.07			-10.42	-10.98

[1] 52 runs.

[2] 62 runs.

Table 14

Bias Associated with Analysis Strategies

Model

Growth model: linear

Treatment effect: interactive (largest effect for lowest status)

Assignment: $w_{t_c} = .20$ $w_{\pi} = .40$ $w_d = .40$

	Distributions for Growth Parameters							
Analysis	1	2	3	4	5	6	7	8
1	-6.13	-6.27	-9.62	-9.83	-4.37	-3.97	-7.42	-7.65
2	-5.38	-5.39	-7.68	-7.84	-5.37	-5.36	-7.61	-7.75
3	-4.94	-4.85	-6.22	-6.29	-5.53	-5.60	-7.66	-7.79
4	-5.46	-5.51	-7.46	-7.61	-5.46	-5.47	-7.68	-7.82
5	-5.30	-5.29	-7.09	-7.21	-5.36	-5.34	-7.61	-7.77
6	-5.15	-5.10	-6.75	-6.82	-5.41	-5.41	-7.62	-7.77
7	-5.24	-5.22	-7.02	-7.17	-5.45	-5.47	-7.63	-7.76
8	-5.09	-5.02	-6.67	-6.77	-5.51	-5.55	-7.64	-7.77
9	-5.07	-5.12	-6.99	-7.20	-5.32	-5.33	-7.50	-7.70
10	-2.55	-2.63	-3.49	-3.97	-2.78	-2.72	-3.63	-3.68
11	-6.92	-7.32	-8.32	-8.69	-5.86	-6.08	-7.52	-7.63
12	-4.94	-4.82	-5.87	-5.77	-5.00	-4.79	-6.48	-6.22
13	-3.29	-2.39	-4.08	-3.32	-4.39	-3.76	-5.59	-4.87
14	-4.92	-4.84	-5.63	-5.45	-5.23	-5.16	-6.67	-6.25
15	-2.24	- .81	-2.71	- .89	-4.43	-3.72	-5.43	-4.11
16	-6.75	-7.07	-8.76	-9.09	-6.26	-6.58	-8.55	-8.91

Table 15

Bias Associated with Analysis Strategies

Model

Growth model: linear

Treatment effect: interactive (largest effect for lowest status)

Assignment: $w_{t_o} = .40$ $w_{\pi} = .20$ $w_d = .40$

Analysis	Distributions for Growth Parameters							
	1	2	3	4	5[1]	6[2]	7[3]	8[4]
1	- .05	- .01	-1.64	-1.75				
2	-2.86	-2.85	- .397	-4.02				
3	-4.82	-4.89	-5.97	-6.05				
4	-4.37	-4.45	-5.49	-5.56				
5	-3.35	-3.38	-4.92	-4.99				
6	-3.93	-4.02	-5.37	-5.47				
7	-3.58	-3.53	-5.07	-5.13				
8	-4.20	-4.2	-5.54	-5.64				
9	-3.21	-3.14	-4.59	-4.65				
10	-1.12	-1.37	-1.83	-1.92				
11	-6.02	-6.69	-6.78	-7.25				
12	-2.76	-2.68	-3.44	-3.19				
13	-2.2-	-1.58	-2.70	-2.02				
14	-3.17	-3.05	-3.49	-3.09				
15	-2.18	- .80	-2.14	- .64				
16	- .47	-4.97	-5.59	-6.10				

[1] 16 runs.

[2] 20 runs.

[3] 55 runs.

[4] 23 runs.

Table 16

Bias Associated with Analysis Strategies

Model

Growth model: linear

Treatment effect: interactive (largest effect for lowest status)

Assignment: $w_{t_o} = -.20$ $w_{\pi} = .20$ $w_d = .60$

	Distributions for Growth Parameters							
Analysis	1	2[1]	3[2]	4	5	6	7	8
1	-5.32		-6.92	-6.84	-7109	-7.14	-8.31	-8.57
2	-2.45		-3.52	-3.48	-2.51	-2.42	-3.35	-3.49
3	- .50		- .52	- .45	-1.78	-1.55	-1.84	-1.82
4	-3.35		-3.67	-3.66	-3.52	-3.49	-3.96	-4.11
5	-2.18		-2.34	-2.31	-2.79	-2.68	-3.19	-3.26
6	-1.62		-1.72	-1.63	-2.56	-2.44	-2.92	-2.98
7	-1.85		-2.01	-1.96	-2.26	-2.12	-2.58	-2.63
8	-1.24		-1.34	-1.23	-2.00	-1.86	-2.28	-2.32
9	-2.25		-2.61	-2.35	-2.62	-2.43	-3.20	-3.26
10	- .90		-1.57	-1.49	-1.21	-1.31	-1.42	-1.65
11	-3.45		-3.74	-3.80	-3.50	-3.44	-3.76	-3.86
12	-2.13		-2.26	-2.31	-2.62	-2.47	-2.82	-2.82
13	-1.22		-1.32	-1.30	-2.14	-1.88	-2.24	-2.11
14	-1.86		-2.01	-2.05	-2.14	-1.96	-2.29	-2.31
15	- .87		-1.00	- .92	-1.63	-1.35	-1.67	-1.57
16	-2.87		-3.29	-3.27	-3.41	-3.20	-3.81	-4.03

[1] 36 runs.

[2] 336 runs.

Appendix B

USER'S GUIDE TO THE COMPUTER PROGRAM

(See Appendix C)

Thomas Jakob Marx

GENERAL DESCRIPTION OF PROGRAM

Input to the Program

For each set of simulations design, the user describes the conditions under which the simulations take place. The maximum and minimum number of simulations is given along with a convergence factor that enables the program to decide whether estimates are stable enough so that it can terminate before the maximum number of simulations is reached.

Information about the sample sizes is also furnished. A maximum total sample size, n_t, and a maximum program sample size, n_p, are specified. From this the computer program determines a sampling fraction. Up to five additional total sample sizes may be requested provided that each one is smaller than the preceding one. In each case, the sampling fraction computed above is applied to fix a program and control sample size.

The user assigns subjects to groups by specifying a set of weights to be used in determining, A_i. If the absolute sum of these weights is less than unity, the difference is used to weight a standardized, normal, random variable.

The user must also specify the conditions under which the growth parameters (onset date, growth notes and birth date) are generated. He must select a distribution for growth parameter--uniform, normal, beta or constant--and set values for the appropriate parameters of each distribution:

Uniform - high and low value

Normal - mean and variance

Beta - alpha (or p) and beta (or g)

Constant - constant value

A functional form for the growth curve is also selected. Options are: linear, quadratic, logistic, and negative exponential.

On the basis of this informaion, the program determines a growth curve for each child and evaluates this curve at the pre- and post-test time points, t_1 and t_2, set by the user.

Each child in the program group has an additional variable called an effect variable. The user has some options regarding its determination. The treatment effect may be constant for all program children. If the effect varies it will be distributed normally. However, by specifying a functional form for the treatment effect rate a user can cause the value of this variable to change according to whatever model of the treatment effect the user desires. Specification of these alternative models requires a program alteration in the EFFECT Subroutine

The final distribution the user must consider is for the Random Component, $R_i(t)$. Although $R_i(t)$ is always distributed normally with a mean of zero, the user must specify a variance.

Related to this is the question of reliability. Since this is a necessary component in several of the analysis strategies considered, it too must be specified. The user can

either use trial runs of the programs to estimate reliability or calculate a population value on the basis of equation 4.15.

Generation of Data Sets

The user will usually request a large number of simulation runs for each set of conditions. (Note: There may be multiple sets of conditions stacked within a single computer run.) Each simulation run produces a data set containing information on a sample of children. In addition to the various growth parameters (onset date, growth rate, and birth date), group membership systematic and observed pre- and post-test scores and age are included. In additoon, for program children, an effect variable representing the benefit of the program to that child is computed.

Basic Statistics

Once a data set is generated, basic statistics are calculated for the total sample and within both program and control groups. The five variables that figure in these statistics are systematic and observed scores at time 1 and time 2 and age. Means, standard deviations, correlation matrices and covariance matrices are also estimated.

Admission of Data Sets

In the final phase of the program sixteen analysis strategies are applied to each data set to estimate the treatment effect. Before subjecting a data set to such an analysis the program examines the data for "reasonableness" according to the following criteria:

1. Pretest difference (observed) must be less than five times the standard error of measurement of the pretest difference;
2. Pre-test variance ratio must be greater than 1/3 and less than 3;
3. Pre-test/post-test correlation must be greater than .1Ø;
4. Index of educational significance must be between Ø and 3ØØ per cent;
5. No correlation involving age and another variable may be below -.2Ø in either group;
6. For the total sample no more than 10% of pre-test or 10% of post-test scores can be Ø or 1ØØ;
7. For the total sample the sums of Ø and 1ØØ scores may be no more than 15% of the total number of scores.

Should any of the criteria listed above not be met, that data set is rejected, no data analysis is undertaken, and the program loops back to begin production of a new data set.

If the number of data sets rejected at any time exceeds 5% of the data sets produced the program terminates the simulation run for that set of conditions and looks for another set of conditions in the input stream. However, a minimum of six data sets must be examined before the 5% barrier is erected.

Determining the Number of Simulations

As mentioned in the input section, the user must specify the minimum and maximum number of simulation runs. When the program hits the minimum, it estimates how many more simulations are necessary to provide acceptable stability to the analyses estimates.

Acceptable means acceptable to the user and she or he must supply a convergence factor deemed reasonable. For example, the user might want error to be no more than .04 and the probability of error exceeding .04 to be no more than .05. On the basis of these two numbers, the convergence factor can be calculated.

The program takes this information and links it to the current estimate of the variance of $\hat{\alpha}$ for each of the sixteen analyses. Using the highest (worst) variance estimate, the number of simulations needed for acceptable covergence is computed. Should this be greater than the maximum number of simulations specified, the run is terminated when the maximum is encountered. Otherwise, the computed number of simulations is performed.

Comparison of Sixteen Analyses

The main question of interst of course, is which analyses estimate the treatment with low bias and good (low) precision. Each technique is applied to the data set and three running figures are tallied: an estimate of bias equal to the true effect minus the estimated effect; the variance of the bias estimate; and the precision of the effect estimates (variance of the estimate plus the bias squared).

When the last simulation run has been finished, these statistics are printed for each of the analyses. In addition, the basic statistics (discussed in the previous section) for the maximum sample sizes are averaged across the simulation runs and printed out. This is also done with the natural effect (average development without the benefit of the program), true effect (average increment to development for children in the program) and the index of educational significance which expresses the true effect as a percentage of the natural effect.

Variable Name Card Format

Card/Cols	Type of Field	Description	Codes
1-8	free	name for time 1 true score	any symbols
10-17	free	name for time 1 observed score	any symbols
19-26	free	name for time 2 true score	any symbols
28-35	free	name for time 2 observed score	any svmbols
37-44	free	name for age at time 1	any symbols

Random Numbers Card Format

1-13	integer	initial random number for this run. (Use 14737222891 for first run)	whole number
15-27	integer	position of initial random number in sequence. (Use 1 for first run) NB: For subsequent runs use the last random and position numbers from the previous run.	whole number

Control Cards Format

page 1 of 4

Card/ Cols	Type of Field	Description	Codes
1/1-4	decimal	maximum number of simulations	whole number
1/5-8	integer	maximum in program group	2-298
1/9-12	integer	maximum sample size first pass	4-3ØØ
1/13-16	integer	sample size second pass (must be less than previous sample size)	4-299
1/17-2Ø	integer	sample size third pass (must be less than previous sample size)	4-298
1/21-24	integer	sample size fourth pass (must be less than previous sample size)	4-297
1/25-28	integer	sample size fifth pass (must be less than previous sample size)	4-296
1/29-32	integer	sample size sixth pass (must be less than previous sample size)	4-295
1/34-39	decimal	program entry time (time 1)	any number
1/41-46	decimal	program exit time (time 2)	any number greater than time 1
1/48-53	decimal	group assignment weight on onset date (absolute sum of weights < = 1)	-1.Ø to 1.Ø
1/55-6Ø	decimal	group assignment weight on growth variable (absolute sum of weights < = 1)	-1.Ø to 1.Ø
1/62-67	decimal	group assignment weight on birth date (absolute sum of weights < = 1)	-1.Ø to 1.Ø

Control Cards Format (continued)

page 2 of 4

Card/ Cols	Type of Field	Description	Codes
2/1	alpha-betic	distribution for onset age	U = uniform N = normal B = beta C = constant
2/2	optional	left parenthesis to improve readability	(
2/3-8	decimal	mean if 2/1 = N,C lowest value if 2/1 = U alpha (or p) if 2/1 = B	any number any number number > Ø
2/9	optional	comma to improve readability	,
2/1Ø-15	decimal	variance if 2/1 = N omit if 2/1 = C highest value if 2/1 = U beta (or q) if 2/1 = B	positive number blank or Ø.Ø any number number > Ø
2/16	optional	right parenthesis to improve readability	)
2/18		distribution for growth variable	same as 2/1
2/19			same as 2/2
2/20-25			same as 2/3-8
2/26			same as 2/9
2/27-32			same as 2/1Ø-15
2/33			same as 2/16
2/35		distribution for birth date	same as 2/1
2/36			same as 2/2
2/37-42			same as 2/3-8
2/43			same as 2/9

Control Cards Format (continued)

page 3 of 4

Card/ Cols	Type of Field	Description	Codes
2/44-49			same as 2/1Ø-15
2/5Ø			same as 2/16
2/8Ø	integer	equation for development rate	1 = linear 2 = quadratic 3 = logistic 4 = negative exponential
3/1	alpha-betic	distribution for first effect variable	N,C
3/2			same as 2/2
3/3-8	decimal	mean	any number
3/9			same as 2/9
3/1Ø-15	decimal	variance if 3/1 = N omit if 3/1 = C	non-negative number blank or Ø.Ø
3/16			same as 2/16
3/4Ø	integer	equation for effect rate in program group	constant = blank or zero high linear = 1 middle linear = 2 low linear = 3
3/41	alpha-betic	distribution of errors at time 1	normal = K (K must be coded)
3/42			same as 2/2
3/43-48	decimal	mean of errors	must be Ø.Ø or blank
3/49			same as 2/9
3/5Ø-55	decimal	variance of errors	number > Ø
3/56			same as 2/16

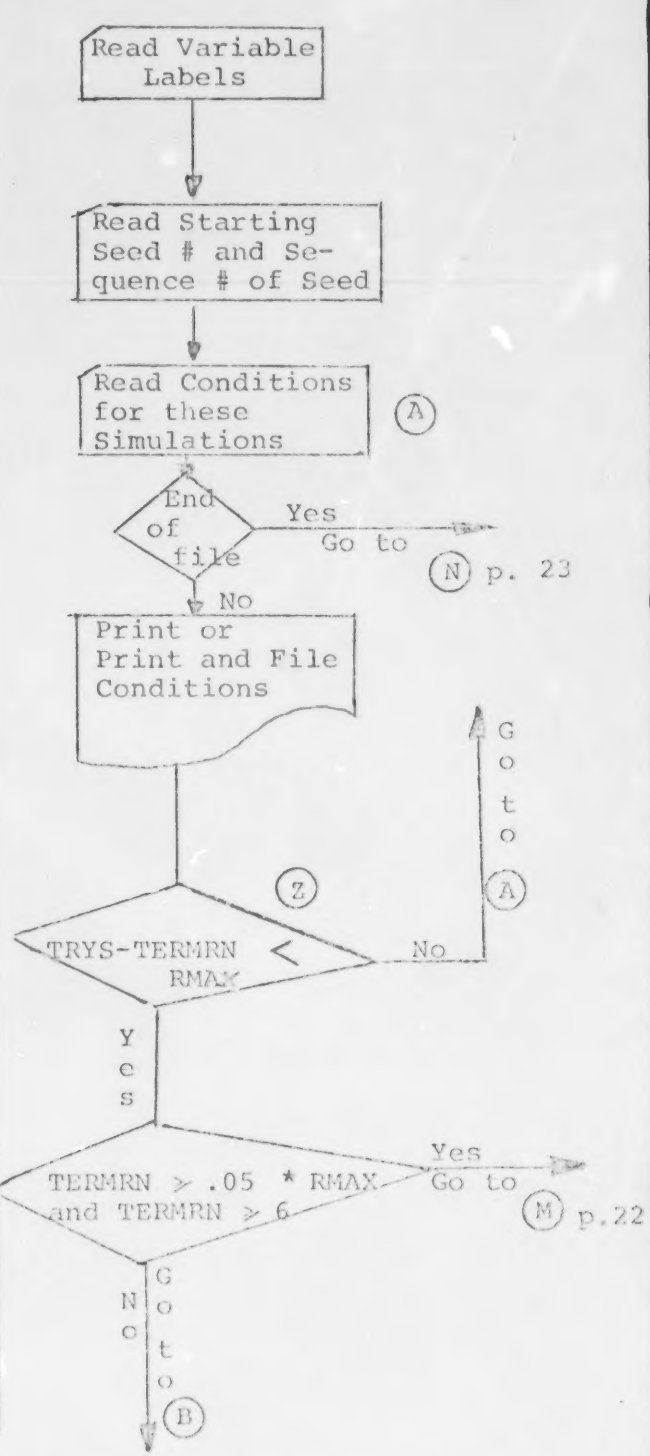
Control Cards Format (continued)

page 4 of 4

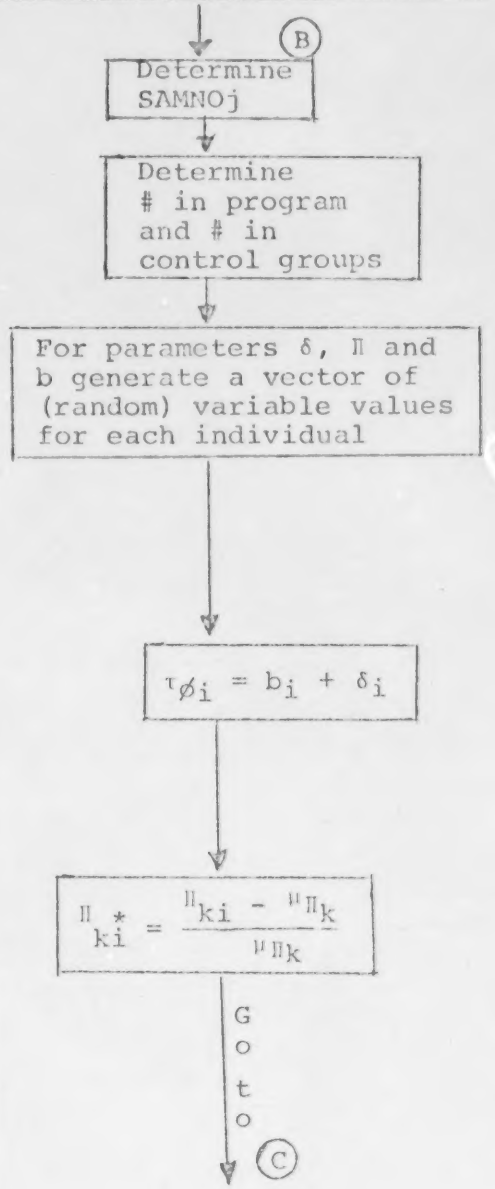
Card/ Cols	Type of Field	Description	Codes
3/58		distribution of errors at time 2	same as 3/41
3/59			same as 2/2
3/6Ø-65			same as 3/43-48
3/66			same as 2/9
3/67-72			same as 3/5Ø-55
3/73			same as 2/16
3/75-78	decimal	reliability of pretest	positive number not greater than 1.Ø
3/79	alpha-betic	print individual values (T = yes for small runs only)	F = no T = yes
3/8Ø	alpha-betic	print summary of each data set (T = yes for preliminary checking)	same as 3/79
4/1-4	integer	minimum number of simulations	positive number over 2
4/5-12	decimal	convergence factor $(Z / \text{tolerance})^2$ to help program decide how many simulations are needed	positive number
4/13-15	integer	logical unit number of tape or disc if user wishes to save printer output. (JCL card should be logical and have unit number that matches this number.)	positive number from 7 to 92 inclusive

Macro Plan of Simulation Program

Notes	Flow Diagram

User supplies variable labels for 5 variables: time 1 growth, time 1 test score (pretest), time 2 growth, time 2 test score (posttest) and age.
 	
Seed # = initial pseudo-random number.
Sequence # = position of seed # in pseudo-random number sequence
 	
The Conditions will be defined as they are used by the program.
 	
The user can request that output be written to tape or disc in addition to being printed.
 	
TRYS = # simulations attempted.
TERMRN = # simulations that failed to meet reasonableness criteria
 	
RMAX = maximum number of simulations for this set of conditions.
	

Macro Plan of Simulation Program (continued)

Notes	Flow Diagram
SAMNO(j) = total in sample j, = 1, 6 and j > j + 1. Supplied as condition. User supplies sampling fraction as condition.	
δ = onset age Π = growth value b = birthdate Each parameter above may be a constant or follow the uniform, normal or beta distribution. User supplies as conditions for δ, Π and b the distribution and chooses appropriate distribution parameters. $\tau\phi_i$ = onset age. (i is the subscript for the individual.)	
$\Pi_{\phi i} = \tau\phi_j$, $\Pi_{1i} = \Pi_i$, $\Pi_{2i} = bi$. $(\Pi^*_{1i}, \Pi^*_{2i}, \Pi^*_{3i})$ is a vector of standard scores for individual i	

Macro Plan of Simulation Program (continued)

Notes	Flow Diagram
W_k = assignment weight between -1 and 1 and $\Sigma\|W_k\| \leq 1$. W_k's are conditions.	(C) $A_i = \sum_{k=\phi}^{2} W_k \, \Pi_{ki}^{*}$
W_3 = weight for random component of A_i	$W_3 = 1 - \Sigma\|W_k\|$
d_i = random variable distributed $N(\phi,1)$.	$A_i = A_i + W_3 d_i$
A_i determines program and control group students. First n_c are control; next n_p are program. $(n = n_c + n_p)$.	Logical sort of students in descending rank on A_i
τ_1 = time 1 or starting time and is a condition. $TEMPl_i$ = growth time at time 1.	$TEMPl_i = \tau_i - \tau\phi_i$
The differential equation form is a condition and may be linear, quadratic, negative exponential or logistic.	Determine: $\frac{dG}{dt} = g\,(\Pi, G; t)$, Go to (D)

Macro Plan of Simulation Program (continued)

Notes

(a) The four differential forms yield analytic solutions symbolized by G_{il}, natural growth at time 1.

(b) R_{il} = error component of G_{il} at time 1.

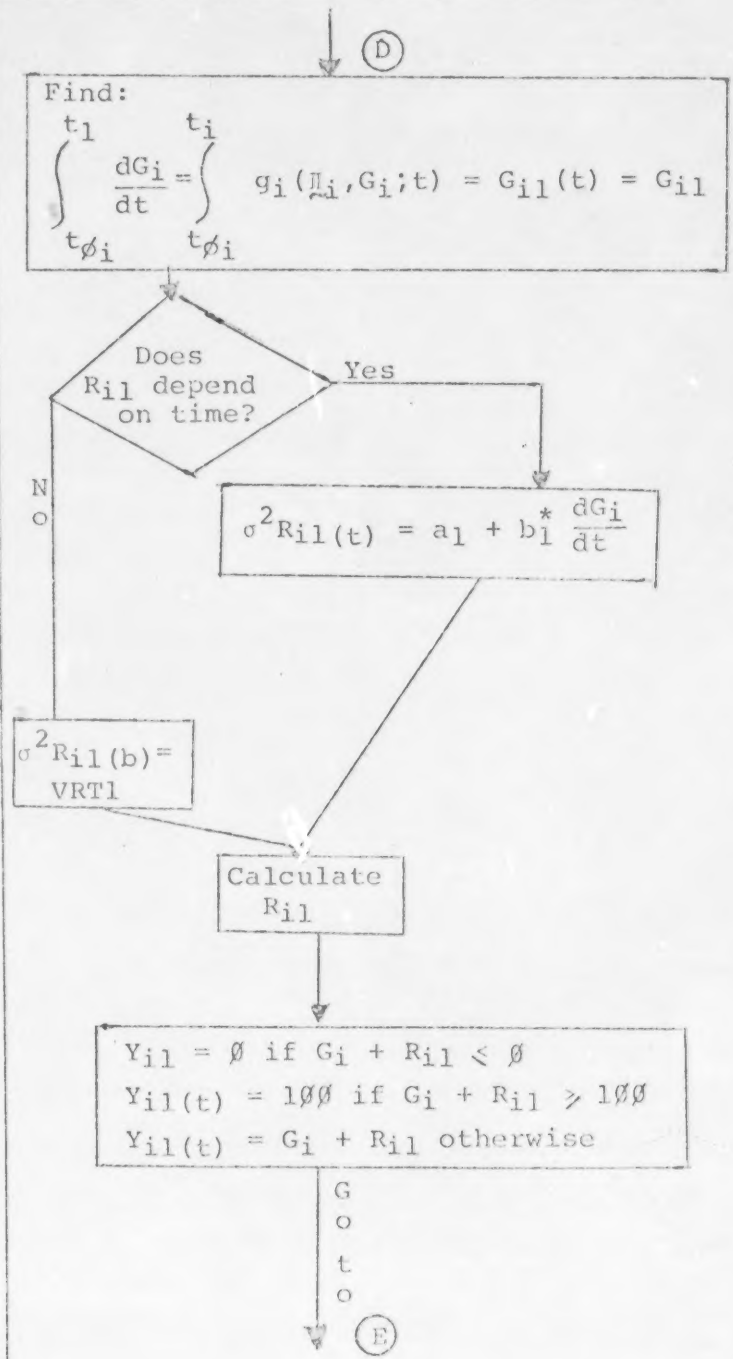
(c) a_1 and b_1 are constants supplied as conditions. var $(R_{il}(t))$ = variance of the error component where E $(R_{il}(t)) = \emptyset$ (E is the expectation.)

(d) VRT1 = constant variance supplied as condition.

(e) R_{il} is distributed $N(\emptyset, \text{var}\ (R_{il}))$

(f) Y_{il} = observed test score at time 1.

Flow Diagram


D
Find:
$\int_{t_{\emptyset i}}^{t_1} \frac{dG_i}{dt} = \int_{t_{\emptyset i}}^{t_i} g_i(\underset{\sim}{\Pi}_i, G_i; t) = G_{il}(t) = G_{il}$
Does R_{il} depend on time?
Yes
No
$\sigma^2 R_{il}(t) = a_1 + b_1^* \frac{dG_i}{dt}$
$\sigma^2 R_{il}(b) =$ VRT1
Calculate R_{il}
$Y_{il} = \emptyset$ if $G_i + R_{il} \leq \emptyset$
$Y_{il}(t) = 100$ if $G_i + R_{il} \geq 100$
$Y_{il}(t) = G_i + R_{il}$ otherwise
Go to
E

Macro Plan of Simulation Program (continued)

Notes	Flow Diagram
	(E)
See (a)	Find: $\int_{t_{\phi i}}^{t_2} \frac{dG_i}{dt} = \int_{t_{\phi i}}^{t_2} g_i(\underset{\sim}{\Pi}_i, G_i; t) = G_{i2}(t) = G_{i2}$
$P_i = \begin{cases} 1 \text{ if student in control group} \\ \phi \text{ if student in program group} \end{cases}$	$P_i = \phi$ if $A_i \geq n_c^{th} A_i$ $= 1$ if $A_i \leq n_c^{th} A_i$
$\underset{\sim}{\varepsilon}$ = vector of treatment (effect) parameters. Distributions and distribution parameters of $\underset{\sim}{\varepsilon}$ are conditions. The form of dI/dt is high, middle or low linear and is also a condition.	Determine: $\frac{dI}{dt} = h(\underset{\sim}{\varepsilon}, G_i; t)$
S_i = systematic growth at time 2 (For controls this is $S_i = G_{i2}$)	$S_i = G_{i2} + P_j \int_{t_1}^{t_2} \frac{dI}{dt}$
	Goto (F)

Macro Plan of Simulation Program (continued)

Notes	Flow Diagram

(g) $NATEFF_S$ = growth from time 1 to time 2 for program children absent a treatment for simulation s.

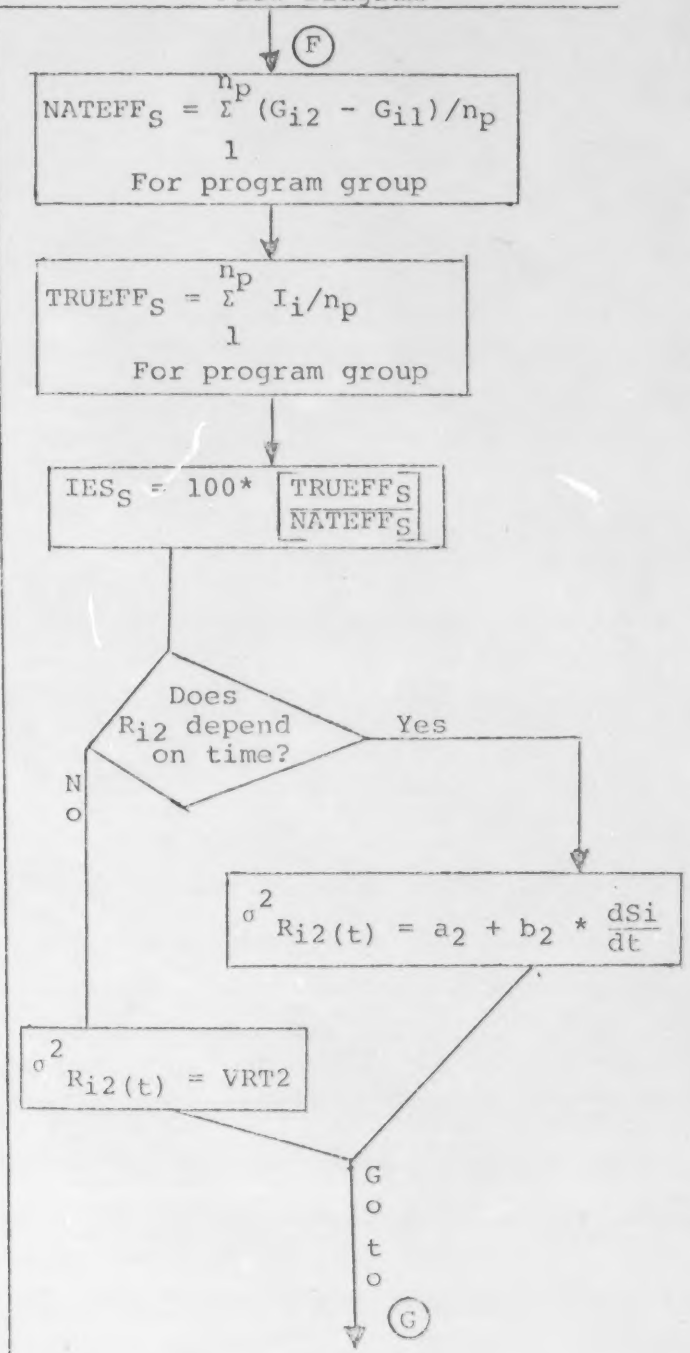
(h) I = treatment effect from time 1 to time 2 for program children.

(i) IES = index of educational significance, a measure of the relative size of the treatment.

see (b)

see (c)

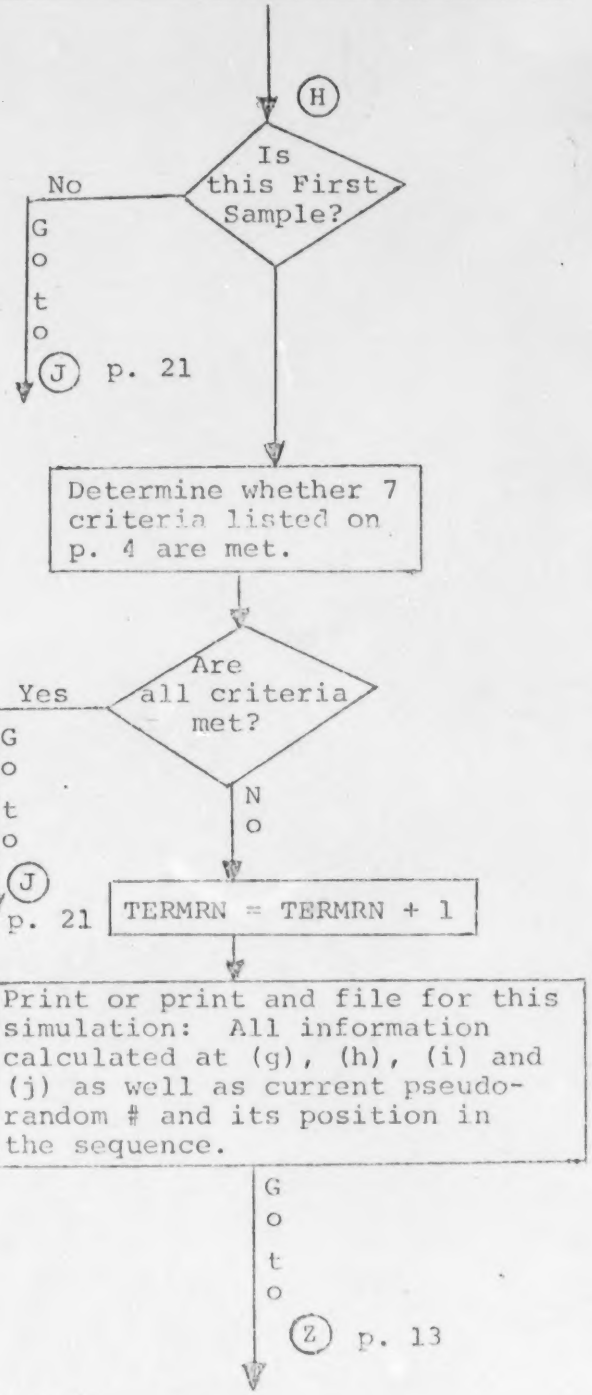
see (d)



Macro Plan of Simulation Program (continued)

Notes	Flow Diagram
	(G)
see (e)	Calculate R_{i2}
see (f)	$Y_{i2}(t) = \emptyset$ if $Si + R_{i2} \leq \emptyset$ $Y_{i2}(t) = 1\emptyset\emptyset$ if $Si + R_{i2} \geq 1\emptyset\emptyset$ $Y_{i2}(t) = Si + R_{i2}$ otherwise
A_i = age at time 1.	$A_i = t_1 - b_i$
(j)	Calculate means, standard deviations, variance-covariance matrix and correlation matrix within total sample, control group and program group for variables G_{i1}, Y_{i1}, Si, Y_{i2} and A_i
	Add statistics of previous box to running totals for these simulations
	Go to (H)

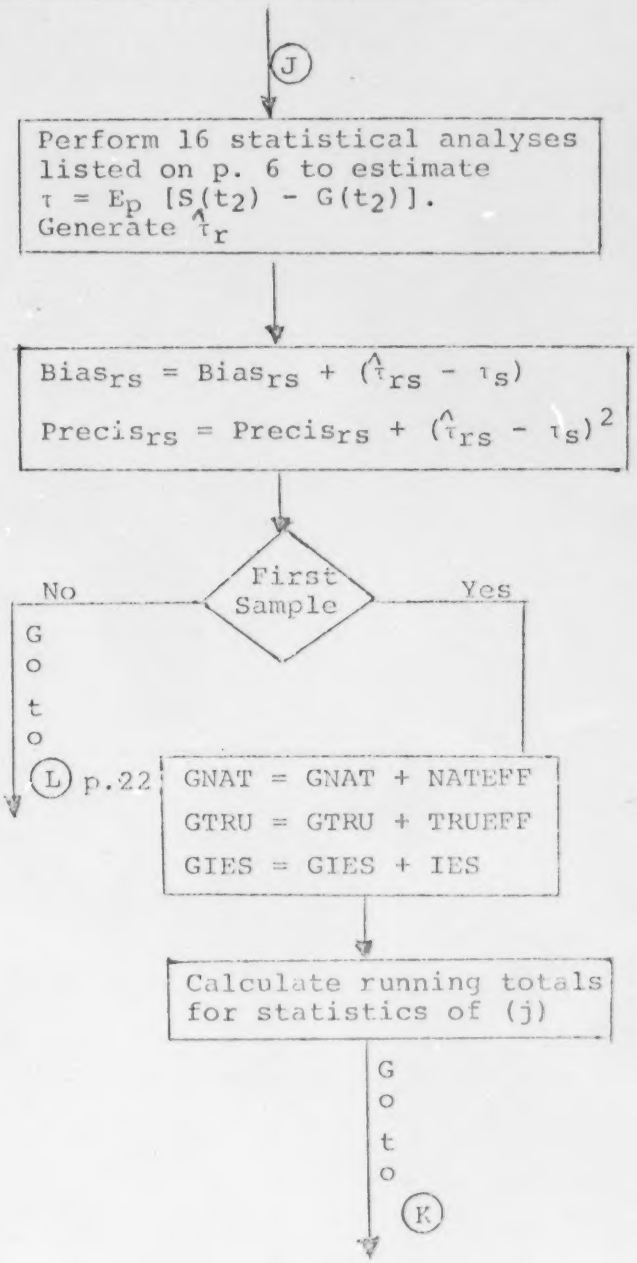
Macro Plan of Simulation Program (continued)

Notes	Flow Diagram
Up to six samples can be used in a simulation, that is SAMNO$_j$ may have a j as large as 6.	
Program checks whether generated data sets are plausible.	
NB: When data set rejected branch back to Z on p. 13 for another possible) simulation.	

H
Is this First Sample?
No
Goto
J p. 21
Determine whether 7 criteria listed on p. 4 are met.
Are all criteria met?
Yes
Goto
J p. 21
No
TERMRN = TERMRN + 1
Print or print and file for this simulation: All information calculated at (g), (h), (i) and (j) as well as current pseudo-random # and its position in the sequence.
Goto
Z p. 13

Macro Plan of Simulation Program (continued)

Notes	Flow Diagram


J
Perform 16 statistical analyses listed on p. 6 to estimate $\tau = E_p [S(t_2) - G(t_2)]$. Generate $\hat{\tau}_r$
$Bias_{rs} = Bias_{rs} + (\hat{\tau}_{rs} - \tau_s)$
$Precis_{rs} = Precis_{rs} + (\hat{\tau}_{rs} - \tau_s)^2$
First Sample
No
Yes
Go to (L) p.22
GNAT = GNAT + NATEFF
GTRU = GTRU + TRUEFF
GIES = GIES + IES
Calculate running totals for statistics of (j)
Go to (K)

τ = true treatment effect
$\hat{\tau}_r$ = treatment estimate based only on observable variables, r = 1,16.

Running bias and precision calculated up through this simulation.

(k) Running totals of natural effect, true effect and IES up through this simulation.

(l)

Macro Plan of Simulation Program (continued)

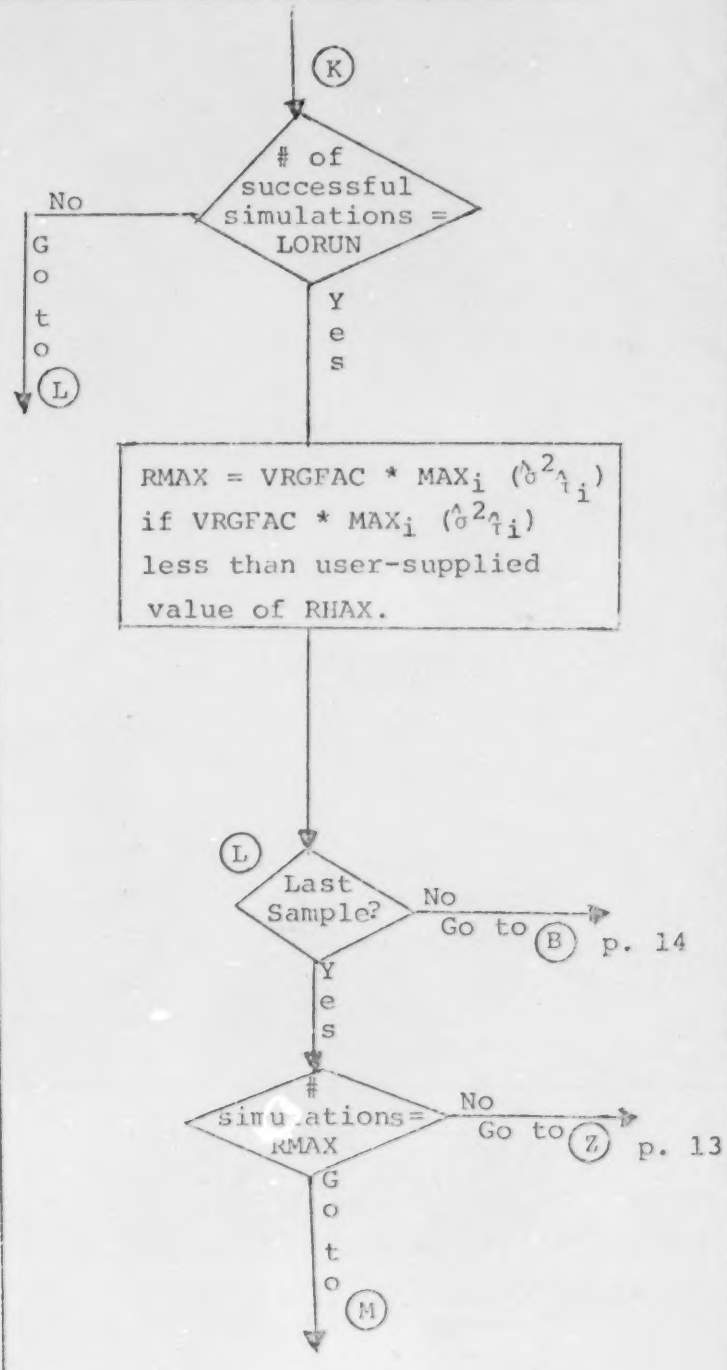
Notes	Flow Diagram

LORUN = minimum # of simulations and is condition.

VRGFAC = $Z^2/(\bar{x}_i - \mu_i')^2$

where Z is distributed N(Ø,1). User supplies VRGFAC as condition after choosing acceptable Z and $(\bar{X}_i - \mu_i')$ values.

$\hat{\sigma}^2_{\hat{\tau}_i}$ = estimated variance of effect estimate from i^{th} analysis after LORUN simulations.



Macro Plan of Simulation Program (continued)

Notes	Flow Diagram
(m)	(M) Divide all statistics of (k) and (l) by # of simulations done to get population approximations
(n) $VRTHAT_i$ = variance of $\hat{\tau}_i$ after last simulation.	For 16 analyses calculate final $BIAS_i$, $PRECIS_i$ and $VRTHAT_i = PRECIS_i - BIAS_i^2$.
	Print or print and file all information from m and n and current pseudo-random number and its position. Go to (A) p. 13
	(N) End of Job
End of job can only be reached after last set of condition cards is read.	

Appendix C

Fortran IV Computer Program for Generating Quasi-Experimental Data in Situations Involving Growth and Examining the Effectiveness of Analysis Strategies.

```
C
C GROWTH(1,J)=ONSET AGE,GROWTH(2,J)=HEREDITY,GROWTH(3,J)=ENVIRONMENT,
C GROWTH(4,J)=BIRTHDATE. 2ND GROWTH(1,J) = BIRTHDATE + ONSET AGE =
C ONSET DATE.
C NB: IF LESS THAN 4 GROWTH VARIABLES ARE USED, THE LAST IS ALWAYS BIRTH
C DATE AND THE FIRST IS ALWAYS ONSET AGE.
C
C
      LOGICAL*1 SCORES,PNTOUT,ADMIT
      INTEGER*2 $SAMNO,$MPARM,$EPARM,CODE(3),DTASET,PAGE,STAR(10)
      REAL IES,NATEFF,DATMAT(300,5),VARNAM(2,5),GTRU(6)
      DOUBLE PRECISION COVCOR(5,5,3),MEANS(5,3),SDS(5,3)
      INTEGER*2 HEX(4),MPARAM,NPARAM,EPARAM,
     $TDIST(4),AT(300),EDIST(2)      ,TEMPAT,SAMPNO,        TMORDS(2)
      INTEGER*2 Z4040,     DSEORM,DEEORM
      INTEGER*4 SEED,CONTRL,PRINT,NSEED,SAMSIZ(6),ROW
      REAL SD(4),MEAN(4),WEIGHT(5),       A(300),R1(2),RS(2),GROWTH(4,
     $300),S1(300),S2(300), R1(300),R2(300),Y1(300),Y2(300),TM(2),
     $TIME(2),EEE(2,300),TREAT(300),SDR1(300),SDR2(300),VALUES(300)
      REAL  TP1(4),TP2(4),EP1(2),EP2(2),DVIATE(300),ESTMAT(16),
     @GROUP(2,3)
     @         ,GEONAM(12),EEONAM(12)
      EQUIVALENCE (EEE(1),DVIATE(1),SDR1(1)),       (EEE(301),SDR2(1))
     $      ,(R1(1),R2(1)),(TM(1),TIME(1))
      EQUIVALENCE(DATMAT(1,1),S1(1)),(DATMAT(1,2),S2(1)),
     $(DATMAT(1,3),Y1(1)),(DATMAT(1,4),Y2(1)),(DATMAT(1,5),VALUES(1))
      COMMON /LOVERC/ LORUN,VRGEAC
      COMMON/KNOCK/RMAX,TERMRN,TRYS,SEX3,DIEOBS,RATVAR,IES,$000(2),
     @$100(2),SUMZAH(2),PRINT,STAR,ADMIT,PNTOUT
      COMMON /STATS/ MEANS,SDS,COVCOR,NT,NC,NP,DATMAT
     @   ,POPMN(5,3),POPSD(5,3),POPCR(5,5,3),AT
      COMMON /RSPREC/BIAS(16,5),PRECIS(16,5),VRTHAT(16,5)
      DATA Z4040/' '/,MPARAM/4/,GROUP/'EVER','YONE','CONT','ROL','PROG',
     @
     @RAM'/,GEONAM/'LINE','AR',' ','QUAD','RATI','C','MOD ','LOGI','STIC
     @','-EXP','ONEN','TIAL'/,EEONAM/'CONS','TANT',' ','HIGH',' LIN',
       'EAR','LOW ','LINE','AR','MID ','LINE','AR'/
      DATA HEX/ZC340,ZC240,ZD540,ZE440/,PAGE/0/
     &,CODE/'D','T','K'/
      DATA $SAMNO,$MPARM,$EPARM/5,4,2/
      ROW = 5
      PRINT = 6
      CONTRL = 5
      NESTIM = 16
      READ(CONTRL,56) VARNAM
56    FORMAT(8(2A4,1X))
      READ(CONTRL,57) SEED,NSEED
57    FORMAT(I13,1X,I13)
   65 READ(CONTRL,58,END=999)RMAX,NPMAX,SAMSIZ,
     $ TM,(WEIGHT(I),I=1,4), (TDIST(I),TP1(I)
     $,TP2(I),I = 1,4),DSEORM,(EDIST(I),EP1(I)
     $                                ,EP2(I),I=1,2),DEEORM,
     $(TMORDS(I),R1(I),RS(I),I=1,2),RHO,SCORES,PNTOUT,LORUN,VRGEAC,
     @ NOISE
58    FORMAT(F4.0 ,7I4,6(1X,F6.0)/
     $     4(A1,2(1X,F6.0),2X),T90,I1/
     $     2(A1,2(1X,F6.0),2X),T40,I1,1X,2(A1,2(1X,F6.0),2X),T75,F4.0,2L1
     $/,I4,F8.0,I3)
      PAGE = PAGE + 1
      NWRITE = 0
413   WRITE (PRINT,64) PAGE
64    FORMAT(1H1,T125,'PAGE',I5)
      WRITE (PRINT,66) RMAX,LORUN,VRGEAC,NPMAX,SAMSIZ,
```

```
     $ TM,(WEIGHT(I),I=1,4), (TDIST(I),TP1(I)
     $,TP2(I),I = 1,4),(GEQNAM(3*(DSEQRM-1) + L),L=1,3)
     $                          ,(EDIST(I),EP1(I)
     $                                    ,EP2(I),I=1,2),
     $(EEQNAM(3*DEEQRM + L),L = 1,3),
     $(TMORDS(I),RI(I),PS(I),I=1,2),PHO,SCORES,PNTOUT,NDISC,SEED,NSEED
  66  FORMAT(1H ,'MAX # OF RUNS IS',T40,F6.0
     8/' ','MIN # OF RUNS IS',T40,I4
     C/' ','(Z**2/TOLERANCE) IS',T40,F10.2
     $                                    /' ','MAXIMUM # IN PROGRAM IS',T4
     10,I6/' ','SAMPLE SIZES ARE',T40,6I6/' ','TIMES 1 & 2 ARE',T40,2F8.
     12/
     2' ','WEIGHTS ARE',T40,4F8.4/' ','GROWTH DISTRIBUTIONS ARE',T40,4(A
     31,'(',F8.4,',',F8.4,')  ')/' ','GROWTH EQUATION IS',T40,3A4
     $                                                          /' ','EF
     4FECT DISTRIBUTIONS ARE',T40,2(A1,'(',F8.4,',',F8.4,')  ')/
     5' ','EFFECT EQUATION IS',T40,3A4
     $                                  /' ','ERROR FUNCTIONS & PARAMETERS
     6ARE',T40,2(A1,'(',F8.4,',',  F8.4,')  ')
     7/' ','RELIABILITY IS',T40,F6.4
     8/' ','INDIVIDUAL SCORES PRINTED?',T40,L1
     9/' ','EACH DATA SET PRINTED?',T40,L1
     D/' ','DISC NUMBER',T40,I4
     A/' ','SEED VALUE & SEED #',T40,2I15)
      NWRITE = NWRITE + 1
      IF(NWRITE .LT. 2) PRINT=PRINT+NDISC
      IF(.NOT.(NWRITE .LT. 2 .AND. NDISC .GT. 0)) GO TO 412
      GO TO 413
 412  PRINT = PRINT - NDISC
C
C TEST FOR MAX RUNS FOR THIS SIMULATION
C
      DTASET = 3
      TERMRN = 0.
      TRYS = 0.
      DO 220 K = 1,3
        DO 220 J = 1,ROW
          POPMN(J,K)=0.
          POPSD(J,K)=0.
          DO 220 I = 1,ROW
 220        POPCR(I,J,K)=0.
      GNAT = 0.
      GIES = 0.
      DO 221 J=1,$SAMNO
        GTRU(J) = 0.0
        DO 221 I = 1,NESTIM
          BIAS(I,J)=0.
          VRTHAT(I,J)=0.
 221      PRECIS(I,J)=0.
 401  IF (TRYS - TERMRN .GE. RMAX) GO TO 65
C
C  HOW MANY SAMPLES?
C
      DO 72 I = 1,$SAMNO
        IF (SAMSIZ (I)) 73,73,72
 72   CONTINUE
      I = I + 1
 73   SAMPNO = I - 1
      NSAMP = 0
 74   NSAMP = NSAMP + 1
      IF (NSAMP .GT. SAMPNO) GO TO 401
      NT = SAMSIZ (NSAMP)
      NP = NT*NPMAX/SAMSIZ(1)
      NC = NT - NP
      NP1ST = NC + 1
      IF (NSAMP .GT. 1) GO TO 763
```

```
C
C         HOW MANY PARAMETERS?
C       DISTRIBUTIONS AND DISTRIBUTION PARAMETERS IN FORM L(NNN,NNN).
C B= BETA, C=CONSTANT, N=NORMAL, U=UNIFORM
C VARIABLE NAMES ARE TDIST,TP1,TP2  WITH SUBSCRIPTS 1 TO MPARAM.
C
      DO 1   I= 1,$MPARM
        IF (TDIST(I).EQ.24040 )GO TO 2
    1   CONTINUE
        I = I + 1
    2 MPARAM = I - 1
C
C         FROM WHICH DISTRIBUTION SHOULD DEVIATES COME ?
C
 10   DO 3  I =  1,MPARAM
        DO 30 K = 1,NPARAM
          IF (TDIST(I) - HEX(K))  30,31,30
   31     GO TO ( 4 , 5 , 6 , 7 ) ,  K
   30     CONTINUE
C
C       GET GROWTH PARAMETERS FOR DISTRIBUTION SPECIFIED.
C
C FIRST  THE  CONSTANT OR 'C' DISTRIBUTION
C
     4 SD(I) =0.0
       MEAN(I)=  TP1(I)
       DO 8 J =  1, NT
     8   GROWTH(I,J) = MEAN(I)
       GO TO 3
C
C SECOND THE BETA DISTRIBUTION
C
     5 SD(I) =SQRT((TP1(I)+TP2(I))/((TP1(I)+TP2(I))**2 * (TP1(I)+TP2(I)+
      $1.0)))
       MEAN(I) = TP1(I)/(TP1(I)+TP2(I))
       CALL BETA(SEED,NSEED,TP1(I),TP2(I),VALUES,NT)
       GO TO 33
C
C THIRD THE NORMAL DISTRIBUTION
C
     6 SD(I) = SQRT(TP2(I))
       MEAN(I) =   TP1(I)
       CALL  NORMAL(SEED,NSEED,MEAN(I),SD(I),VALUES,NT,1)
       GO TO 33
C
C FOURTH THE UNIFORM DISTRIBUTION
C
     7 SD(I) = SQRT((TP2(I) - TP1(I))**2 / 12.0)
       MEAN(I) = (TP1(I) + TP2(I)) / 2.0
       CALL UNIF (SEED,NSEED,TP1(I),TP2(I),VALUES,NT)
    33 DO 34  J=  1,NT
    34   GROWTH(I,J) = VALUES(J)
     3 CONTINUE
       MEAN(1)=MEAN(1)+MEAN(MPARAM)
       SD(1)=SD(1)+SD(MPARAM)
       IF(.NOT.SCORES) GO TO 763
       DO 47 I = 1,MPARAM
       WRITE(PRINT,200) I
   200 FORMAT (1H0, T60,'GROWTH ',I1)
       WRITE(PRINT,201) (GROWTH(I,J),J= 1,NT)
   201 FORMAT(1H , 10F13.3)
    47 CONTINUE
C
C ONSET DATE = ONSET AGE + BIRTH DATE.
C
   763 DO 48 J = 1,NT
```

```
      IF(NSAMP .EQ. 2) GROWTH(1,J)=TM(2)-GROWTH(1,J)
 48   IF(NSAMP .EQ. 1) GROWTH(1,J)=GROWTH(1,J)+GROWTH(MPARAM,J)
      IF(.NOT.SCORES) GO TO 600
      WRITE(PRINT,501)
 501  FORMAT(1HO,T60,'GROWTH 1 TRANSFORM')
      WRITE (PRINT,201)(GROWTH(1,J),J=1,NT)
C
C          WEIGHT ON ERROR TERM
C RANGE IS -1 TO 1; ABSOLUTE SUM OF WEIGHTS = 1.
C
600   MNO=MPARAM-1
      TWT  = 0.0
      DO 12 I = 1, MNO
   12   TWT = ABS(WEIGHT(I)) + TWT
      WEIGHT(MNO+1)= 1.0 - TWT
C
C          ASSIGN  N(0,1) ERROR TERM FOR SAMPLE ASSIGNMENT VARIABLE.
C
   90 CALL NORMAL(SEED,NSEED,0.0,1.0,DVIATE,NT,1)
C
C
C COMPUTE ASSIGNMENT  VARIABLE VALUE FOR EACH SUBJECT.
C
      DO 17  J = 1,NT
        A(J)=0.0
      DO 18 I=1,MNO
          IF(TDIST(I) .EQ.HEX(1)) GO TO 18
          A(J)=((GROWTH(I,J)-MEAN(I))/ SD(I))*WEIGHT(I) + A(J)
   18     CONTINUE
   17  A(J)=DVIATE(J)*WEIGHT(MNO+1) + A(J)
      IF(.NOT.SCORES) GO TO 601
      WRITE(PRINT,502) NSAMP
 502  FORMAT(1HO,T60,'A BEFORE SORT',I1)
      WRITE(PRINT,201)(A(J),J=1,NT)
C
C INITIALIZE ADDRESSES IN AT
C
 601  DO 11 J = 1,NT
 11     AT(J) = J
C
C          RANK ASSIGNMENT VARIABLES IN DESCENDING ORDER
C
      CALL FIZZ(A,AT,NT)
      IF(.NOT.SCORES) GO TO 602
      WRITE (PRINT,503)NSAMP
 503  FORMAT(1HO,T60,'A AFTER SORT',I1)
      WRITE (PRINT,201)(A(AT(J)),J=1,NT)
 602  IF (NSAMP .GT. 1) GO TO 39
C
C CREATE THE T1 - T0I VARIBLE.
C
      DO 603 J=1,NT
 603  GROWTH(1,J)=TM(1)-GROWTH(1,J)
C
C CALL FUNCTION  THAT RETURNS,FOR ALL SUBJECTS, A VALUE OF  S1   AT TIME 1.
C DIFFERENT DIFFERENTIAL EQUATIONS  IN DSCUM:
C  1 = LINEAR
C  2 = QUADRATIC
C  3 = LOGISTIC
C  4 = NEGATIVE EXPONENTIAL
C    THE VARIABLE CONTAINING THESE FORMS IS  DSFORM
C
C
      CALL SOLVE(       NT,DSFORM,GROWTH, S1       )
      IF(TWORDS(1) .EQ. CODE(1))
     $                        CALL VARR(NT,S1,      RI(1),RS(1),
```

```
     $ DSEQRM,GROWTH,SDR1,&25)
      IF(TMORDS(1) .EQ. CODE(2))
     L                    CALL VARRI(NT,      RI(1),RS(2),  GROWTH,SDR1,
     $&25)
      DO 308 I = 1,NT
 308  SDR1(I) = RS(1)
 25   DO 310 I = 1,NT
 310    SDR1(I) = SQRT(SDR1(I))
      CALL NORM1  (SEED,NSEED,0.0,SDR1,R1,NT,1)
      $000(1) = 0.
      $100(1) = 0.
      DO 39 J=1,NT
        Y1(J) =  S1(J) + R1(J)
        IF((Y1(J) - AINT(Y1(J)))- 0.5) 38,37,37
  37    Y1(J) =  Y1(J) + 1.0
  38    Y1(J) =  AINT( Y1(J))
      IF(0..LT.Y1(J).AND.Y1(J).LT.100.)GO TO 39
      IF(100..LE.Y1(J))GO TO 42
      Y1(J)=0.
      $000(1)=$000(1) + 1.
      GO TO 39
  42  Y1(J) = 100.
      $100(1) = $100(1) + 1.
  39  CONTINUE
C
C RANK C & P ON PRETEST
C
      DO 20 K=1,NT
        J = AT(K)
        IF (K - NC) 22,22,21
 21      A(J) = Y1(J)
        GO TO 20
 22      A(J) = Y1(J) + 101.
 20      CONTINUE
      CALL FIZZ(A,AT,NT)
      NTIME = 1
      IF(.NOT.SCORES) GO TO 210
      WRITE(PRINT 504) NTIME
 504  FORMAT(1HO,T60,'PRETEST+101',I2)
      WRITE(PRINT,201)(A(AT(J)),J=1,NT)
      WRITE(PRINT,204) NTIME
  204 FORMAT(1HO, T60,'S',I1)
      WRITE(PRINT,201)(S1(AT(J)),J= 1,NT)
      WRITE(PRINT,205) NTIME
  205 FORMAT(1HO,T60,'R',I1)
      WRITE(PRINT,201)(R1(AT(J)),J= 1,NT)
      WRITE(PRINT,206)NTIME
  206 FORMAT(1HO,T60,'Y',I1)
      WRITE(PRINT,201)(Y1(AT(J)),J = 1,NT)
 210  IF (NSAMP .GT. 1) GO TO 96
      DO 49 J = 1,NT
 49   GROWTH(1,J) = GROWTH(1,J) + TM(2) - TM(1)
C
C CALCULATE S2   FOR EVERYONE AT TIME 2.
C
      CALL  SOLVE(NT,DSEQRM,GROWTH,S2)
C
C MAKE EFFECT VARIANCES INTO SD'S & FIND # OF EFFECTS.
C
      DO 81 I =  1, $EPARM
        IF (EDIST(I) .EQ.24040) GO TO 83
   81   CONTINUE
        I = I + 1
   83 EPARAM = I - 1
  96   IF (.NOT.SCORES) GO TO 211
       WRITE(PRINT,203) NSAMP
```

```
 203 FORMAT(1HO,T60,'ASSIGNMENT',I2)
     WRITE(PRINT,201)(A(AT(J)),J = 1,NT)
 211 DO 85 I= 1,EPARAM
     IF(ABS(EP2(I) - 0.0) .LT. 1E-6) GO TO 301
       SD (I) =  SQRT(EP2(I))
       CALL NORMAL(SEED,NSEED,EP1(I),SD (I),VALUES,NT,NP1ST)
       DO 86 K = NP1ST,NT
         J = AT(K)
  86     EFF(I,J) = VALUES(K)
     GO TO 85
 301 DO 302 K = NP1ST,NT
       J = AT(K)
 302   EFF(I,J) = EP1(I)
  85   CONTINUE
     IF (.NOT.SCORES) GO TO 212
     DO 88 I= 1,EPARAM
       WRITE(PRINT,207) I
 207   FORMAT(1HO, T60,'EFFECT ',I1)
       WRITE(PRINT,201)(EFF(I,AT(J)),J= NP1ST,NT)
  88   CONTINUE
C
C CALCULATE TREATMENT EFFECTS IF THEY DEPEND ON THE INDIVIDUAL.
C
 212 IF (DEFORM .EQ. 0) GO TO 279
     IF (NSAMP .GT. 1)
    +CALL EFFECT(NT,NP1ST,DEFORM,AT,TREAT,EFF,GROWTH,TIME,&75)
     I = 0
 281 DO 280 K = NP1ST,NT
       J = AT(K)
 280   GROWTH(1,J) = TIME(2) - GROWTH(1,J)
     I = I + 1
     IF (I .GT.  ) GO TO 75
     CALL EFFECT,NT,NP1ST,DEFORM,AT,TREAT,EFF,GROWTH,TIME,&281)
 279 DO 63 K = NP1ST,NT
       J = AT(K)
  63   TREAT(J) = EFF(1,J)* (TM(2)-TM(1))
  75 IF (.NOT. SCORES) GO TO 213
     WRITE(PRINT,303)
 303 FORMAT(1HO,T60,'TREATMENT')
     WRITE(PRINT,201)(TREAT(AT(J)),J=NP1ST,NT)
C
C CALCULATE NATURAL EFFECT, THE DENOMINATOR OF THE IES.
C
 213 NATEFF = 0.0
     DO 320 K = NP1ST,NT
       J = AT(K)
 320   NATEFF = NATEFF + S2(J) - S1(J)
C
C CALCULATE TRUEFF, THE NUMERATOR OF THE IES.
C
     TRUEFF = 0.0
     DO 91  K = NP1ST,NT
       J = AT(K)
     S2(J) = S2(J) + TREAT(J)
  91 TRUEFF = TRUEFF + TREAT(J)
     IES = (TRUEFF / NATEFF )  * 100.0
     TRUEFF = TRUEFF /FLOAT(NP)
     NATEFF = NATEFF / FLOAT(NP)
C
C
C SAMPLE FROM THE ERRORS R2(T)
C
      IF(TMORDS(2) .EQ. CODE(1))
     $                          CALL VARR(NT,S2      ,R1(2),RS(2),
     $DSEQRM,GROWTH,SDR2,&105)
      IF(TMORDS(2) .EQ. CODE(2))
```

```
     $                    CALL VARR1(NT      ,R1(2),RS(2),GROWTH,
     $SDR2,&105)
      DO 309 I = 1,NT
  309 SDR2(I) = RS(2)
  105 DO 312 I = 1,NT
  312   SDR2(I) = SQRT(SDR2(I))
      CALL NORM1 (SEED,NSEED,0.0,SDR2,R2,NT,1)
C
C NOW COMPUTE  OBSERVEDS AT TIME 2.
C
      $000(2)=0.
      $100(2)=0.
      DO 112 J = 1,NT
        Y2(J) =  S2(J) + R2(J)
        IF((Y2(J)- AINT(Y2(J)))- 0.5) 110,111,111
  111   Y2(J) = Y2(J) + 1.0
  110   Y2(J) =  AINT(Y2(J))
        IF(0. .LT. Y2(J) .AND. Y2(J) .LT. 100.) GO TO 112
        IF (100. .LE. Y2(J)) GO TO 109
        Y2(J) = 0.
        $000(2) = $000(2) + 1.
        GO TO 112
  109    Y2(J) = 100.
        $100(2) = $100(2) + 1.
  112 CONTINUE
      NTIME=2
      IF (.NOT. SCORES) GO TO 214
      WRITE(PRINT,204)NTIME
      WRITE(PRINT,201)(S2(AT(J)),J = 1,NT)
      WRITE(PRINT,205)NTIME
      WRITE(PRINT,201)(R2(AT(J)),J = 1,NT)
      WRITE(PRINT,206) NTIME
      WRITE(PRINT,201)(Y2(AT(J)),J=1,NT)
  214 DO 337 J = 1,NT
  337   VALUES(J) = TM(1) - GROWTH(MPARAM,J)
      K = 0
  370 K = K + 1
      IF (K .NE. 1) GO TO 375
      NSTART = 1
      NFINIS = NT
      GO TO 372
  375 IF (K.NE.2) GO TO 385
      NFINIS = NC
      GO TO 372
  385 NSTART = NC + 1
      NFINIS = NT
  372 DO 321 I = 1,ROW
        MEANS(I,K) = 0.
        SDS(I,K) = 0.
        DO 321 J = 1, ROW
  321     COVCOR(I,J,K) = 0.0
      CALL ESTATS        (NSTART,NFINIS,K,ROW)
      IF (K .LT. 3) GO TO 370
C
C ONLY CALL NOKNOK FOR MAXIMUM N IN SAMPLE
C
      IF (NSAMP .EQ. 1) CALL NOKNOK(&401,&65,PAGE,NDISC)
      IF (.NOT. ADMIT)
     &             CALL PRINTR(PAGE,DTASET,          ROW,TRUEFF,NATEFF,
     %     NSAMP,ESTMAT,VARNAM,GROUP,SEED,NSEED,NESTIM,&401,NDISC)
      CALL TRUBLU (RHO,TM,              NSAMP,SAMPNO,TRUEFF,NPLST,
     @     A,NESTIM,      GNAT,GTRU,GIES,SCORES,SAMSIZ,
     & PAGE,GROUP,VARNAM,SEED,NSEED,ROW,NATEFF,ESTMAT,DTASET,NDISC)
      IF (NSAMP .EQ. SAMPNO) GO TO 74
      DO 382 K = NPLST,NT
        J = AT(K)
```

```
382     S2(J) = S2(J) - TREAT(J)
      GO TO 74
999   STOP
      END
```

```
      SUBROUTINE BETA(SEED,NSEED,TP1,TP2,VALUES,NT)
      DIMENSION IP(2),X2(2),VALUES(1)
      INTEGER PLANT,SEED
      DATA DEL / 1.0E-5 /
      DO 113 L  = 1,NT
      IP(1) = TP1 + 1E-5
      IP(2) = TP2 + 1E-5
      X2(1) = 0.0
      X2(2) = 0.0
      DO 260 I = 1,2
        K = IP(I)/2
        IF ( K .LE. 0) GO TO 240
        DO 220 J = 1,K
200       PLANT = SEED * 65539
          IF (PLANT) 5,6,6
  5       PLANT = PLANT + 2147483647 + 1
  6       U     = PLANT
          U = U *  .4656613E-9
          SEED = PLANT
          NSEED  = NSEED + 1
          IF (U  .LE. DEL ) GO TO 200
          X2(I) =    X2(I) - ALOG(U)
220       CONTINUE
        X2(I) =  2*X2(I)
240     IF   (IP(I) .LE.  2*K) GO TO 260
        DO   50 M = 1,12
            PLANT = SEED * 65539
            IF (PLANT) 7,8,8
  7         PLANT = PLANT + 2147483647 + 1
  8         ZVAL =  PLANT
            ZVAL =  .4656613E-9 * ZVAL
            SEED = PLANT
            NSEED = NSEED + 1
 50         X2(I)  =  X2(I) +  ZVAL
260     CONTINUE
      VALUES(L) =  X2(1) / ( X2(1) + X2(2))
113   CONTINUE
      RETURN
      END
```

```
      SUBROUTINE NORMAL(SEED,NSEED,MEAN,SD,/VALUES/,NVAL,NSTART)
      DIMENSION VALUES(1),SDVEC(1)
      INTEGER SEED,PLANT
      REAL MEAN
      DO 60 J = NSTART,NVAL
        A  =  0.0
        DO 50 I  =  1,12
        PLANT =  SEED * 65539
        IF  (PLANT)  5,6,6
    5   PLANT =  PLANT + 2147483647 + 1
    6   VALUES(J) = PLANT
        VALUES(J) =  .4656613E-9*VALUES(J)
        SEED = PLANT
        NSEED = NSEED + 1
   50   A = A + VALUES(J)
   60   VALUES(J) = (A - 6.0)*SD + MEAN
        RETURN
      ENTRY NORM1  (SEED,NSEED,MEAN,/SDVEC/,/VALUES/,NVAL,NSTART)
      DO 80 J = NSTART,NVAL
        A  =  0.0
        DO 70 I  =  1,12
        PLANT =  SEED * 65539
        IF  (PLANT)  7,8,8
    7   PLANT =  PLANT + 2147483647 + 1
    8   VALUES(J) = PLANT
        VALUES(J) =  .4656613E-9*VALUES(J)
        SEED = PLANT
        NSEED = NSEED + 1
   70   A = A + VALUES(J)
   80 VALUES(J) = (A - 6.0)*SDVEC(J) + MEAN
      RETURN
        END
```

```
      SUBROUTINE UNIF(SEED,NSEED,LOW,HIGH,/VALUES/,NVAL )
      DIMENSION VALUES(1)
      INTEGER SEED,PLANT
      REAL LOW,HIGH
      DO 60 J = 1,NVAL
        PLANT = SEED*65539
        IF (PLANT) 5,6,6
    5   PLANT = PLANT + 2147483647 + 1
    6   VALUES(J) = PLANT
        VALUES(J) = .4656613E-9*VALUES(J)
      VALUES(J) = LOW*(1. - VALUES(J)) + HIGH*VALUES(J)
        SEED =  PLANT
   60   NSEED =  NSEED   + 1
      RETURN
      END
```

```
      SUBROUTINE SOLVE (NT,DSFORM,/GROWTH/,/S/)
      INTEGER*2 DSFORM
      DIMENSION  GROWTH(4,NT), S(1)
      GO TO (10,20,30,40 ), DSFORM
   10 DO 15 J = 1,NT
   15   S(J) = GROWTH(2,J) * GROWTH(1,J)
      RETURN
   20 DO 25  J=1,NT
   25   S(J) =  GROWTH(2,J)*GROWTH(1,J)  +
     $          GROWTH(3,J)*GROWTH(1,J)**2
      RETURN
   30 DO 35  J=1,NT
   35   S(J)=100.
     $       /(1.+EXP(-GROWTH(2,J)*GROWTH(1,J)))
      RETURN
   40 DO 45 J= 1,NT
   45 S(J)=100. - 100.*EXP( -GROWTH(2,J)*GROWTH(1,J))
      RETURN
      END
```

```
      SUBROUTINE VARR(NT,S     ,RICPT,RSLPE,DSFORM,/GROWTH/,/SDR/,*)
      INTEGER*2 DSFORM
      DIMENSION S(1),GROWTH(4,NT),SDR(1)
C
C CALCULATE  DS/DT.
C
      DO 65  J =  1, NT
      GO TO (10,20,30,40) , DSFORM
   10 DSBYDT =  GROWTH(2,J)
      GO TO 65
   20 DSBYDT = GROWTH(2,J) + GROWTH(2,J)*GROWTH(1,J)
      GO TO 65
   30 DSBYDT = GROWTH(3,J)*(S(J)+ GROWTH(2,J))*(1.0 - S(J))
      GO TO 65
   40 DSBYDT = (1.0 - S(J)) * GROWTH(2,J)
C*****************************************************************************
C TOM MADE DSBYDT ABS(DSBYDT) TO PREVENT NEGATIVE SQUARE ROOTS FROM           *
C OCCURING.                                                                   *
C*****************************************************************************
   65   SDR(J) =  SQRT(RICPT + RSLPE * ABS(DSBYDT))
      RETURN 1
      ENTRY VARR1(NT     ,RICPT,RSLPE,/GROWTH/,/SDR/,*)
      DO 75 J = 1,NT
   75   SDR(J) = SQRT(RICPT +  RSLPE * GROWTH(1,J) )
      RETURN 1
      ENTRY VARR2(NT,RICPT)
      RICPT = SQRT(RICPT)
      DO 85 J = 1,NT
   85   SDR(J) = RICPT
      RETURN
      END
```

```
      SUBROUTINE EFFECT(NT,NPIST,DEFORM,AT,TREAT,EFF,GROWTH,TIME,*)
      DIMENSION EFF(2,NT),TREAT(1),GROWTH(4,NT),TIME(2)
      INTEGER*2  AT(1),DEFORM
        GO TO (10,20,30) ,    DEFORM
   10 DO 15 K =NPIST, NT
        J = AT(K)
   15   TREAT(J)=EFF(1,J)*GROWTH(2,J)*((TIME(2)**2-TIME(1)**2)/2.-
     &  GROWTH(1,J)*(TIME(2)-TIME(1)))
        RETURN 1
   20 DO    25  K = NPIST,NT
        J = AT(K)
   25   TREAT(J) = 100.*EFF(1,J)*(TIME(2)-TIME(1))
     $           -EFF(1,J)*GROWTH(2,J)*((TIME(2)**2-TIME(1)**2)/2.-
     &  GROWTH(1,J)*(TIME(2)-TIME(1)))
        RETURN 1
   30 DO  35    K = NPIST,NT
        J = AT(K)
   35   TREAT(J) =
     $      100.*EFF(1,J)*GROWTH(2,J)*((TIME(2)**2-TIME(1)**2)/2.-
     &  GROWTH(1,J)*(TIME(2)-TIME(1)))
     %  -EFF(1,J)*GROWTH(2,J)**2*((TIME(2)**3-TIME(1)**3)/3.-
     <  GROWTH(1,J)*(TIME(2)**2-TIME(1)**2)-GROWTH(1,J)**2*(TIME(2)-
     ?  TIME(1)))
        RETURN 1
        END


      SUBROUTINE ESTATS (NSTART,NFINIS,L,M)
      DIMENSION X(300,5)
      DOUBLE PRECISION XQ(5,5,3),XBAR(5,3),SD(5,3)
      INTEGER*2 AT(300)
      COMMON /STATS/XBAR,SD,XQ,NT,NC,NP,X,
     @    POPMN(5,3),POPSD(5,3),POPCR(5,5,3),AT
      DO 2 J=1,M
      DO 3 II=NSTART,NFINIS
      I = AT(II)
    3 XBAR(J,L) = XBAR(J,L) + X(I,J)
      XBAR(J,L) = XBAR(J,L) / (NFINIS - NSTART + 1)
      DO 2 II= NSTART,NFINIS
      I = AT(II)
    2 X(I,J) = X(I,J) - XBAR(J,L)
      DO 5 I=1,M
      DO 6 J=1,I
      COVAR = 0.0
      VAR1 = 0.0
      VAR2 = 0.0
      DO 10 II = NSTART,NFINIS
      K = AT(II)
      COVAR = COVAR + X(K,I) * X(K,J)
      VAR1 = VAR1 + X(K,I)**2
      VAR2 = VAR2 + X(K,J)**2
   10 CONTINUE
      IF (ABS(VAR1*VAR2)-1E-5)12,13,13
   12 XQ(J,I,L) = 0.
      GO TO 14
   13 XQ(J,I,L) = COVAR/SQRT(VAR1*VAR2)
   14 XQ(I,J,L) = COVAR/ (NFINIS - NSTART)
    6 CONTINUE
      SD(I,L)=DSQRT(XQ(I,I,L))
    5 CONTINUE
      DO 8 J = 1,M
      DO 8 II = NSTART,NFINIS
      I = AT(II)
    8 X(I,J) = X(I,J) + XBAR(J,L)
      RETURN
      END
```

```
      SUBROUTINE FIZZ (DATA,AT,N)
      REAL DATA(300)
      INTEGER*2 AT(300)
      LIMIT = N
      LAST = 1
      KEY = -1
      SIGN = -1.0
100   J = LAST
      ISTOP = LIMIT
      LIMIT = LAST
      LAST = 0
      KEY = -KEY
      SIGN = -SIGN
110   I = J
      IF (I .EQ. ISTOP ) GO TO 130
      J = I + KEY
      LI = AT(I)
      LJ = AT(J)
      TEST = SIGN * ( DATA(LJ) - DATA(LI) )
      IF (TEST) 110,110,120
120   LAST = I
      AT(I) = LJ
      AT(J) = LI
      GO TO 110
130   IF ( LAST ) 140, 140, 100
140   RETURN
      END


      SUBROUTINE NOKNOK(*,*,PAGE,NDISC)
      COMMON/KNOCK/RMAX, ERMRN,TRYS,SEX3,DIFOBS,RATVAR,IES,$000(2),
     @$100(2),SUMZAH(2),PRINT,STAR,ADMIT,PNTOUT
      INTEGER PRINT
      REAL IES
      INTEGER*2 STAR(10),TESTS,STR,AT(300),BLANK,PAGE
      LOGICAL*1 PNTOUT,ADMIT
      REAL DATMAT(300,5)
      DOUBLE PRECISION COVCOR(5,5,3),MEANS(5,3),SDS(5,3)
      COMMON /STATS/ MEANS,SDS,COVCOR,NT,NC,NP,DATMAT
     @   ,POPMN(5,3),POPSD(5,3),POPCR(5,5,3),AT
      DATA BLANK,STR,TESTS/' ','*',10/
      ADMIT = .TRUE.
      TRYS = TRYS + 1.
      DO 88 I = 1,TESTS
 88     STAR(I) = BLANK
C
C IS |YIP. - YIC.| LE 5*SE ?
C
      DIFOBS=DABS(MEANS(3,3)-MEANS(3,2))
      SEX3=5*DSQRT(COVCOR(3,3,3)/NP+COVCOR(3,3,2)/NC)
      IF (DIFOBS .LE. SEX3) GO TO 1
      STAR(1)= STR
C
C 1/3 LE V(YIP.)/V(YIC.) LE  3
C
  1   RATVAR = COVCOR(3,3,3) / COVCOR(3,3,2)
      IF ( 1./3. .LE. RATVAR .AND. RATVAR .LE. 3.) GO TO 2
      STAR(2) = STR
C
C .1 LE R(PRE P,POST P) LE .95
C
  2   IF (.1 .LE. COVCOR(3,4,3) .AND. COVCOR(3,4,3) .LE. .95) GO TO 3
```

```
      STAR(3)=STR
C
C .1 LE R(PRE C,POST C) LE  .95
C
 3    IF (.1 .LE. COVCOR(3,4,2) .AND. COVCOR(3,4,2) .LE. .95)GO TO 4
      STAR(4) = STR
C
C % LT IES LT 300%
C
 4    IF(0..LT. IES .AND. IES .LT. 300.) GO TO 5
      STAR(5) = STR
C
C ALL CORRELATIONS INVOLVING AGE ARE GT -.20 IN P
C
 5    DO 30 I=1,4
      IF(COVCOR(I,5,3) .LE. -.20) GO TO 25
 30     CONTINUE
      GO TO 6
 25   STAR(6) = STR
C
C ALL CORRELATIONS INVOLVING AGE ARE GT -.20 IN C
C
 6    DO 40 I=1,4
        IF (COVCOR(I,5,2) .LE. -.20) GO TO 35
 40     CONTINUE
      GO TO 7
 35   STAR(7)=STR
C
C # OF ZEROES LT 10% IN Y1 & Y2
C
 7    DO 50 I=1,2
        IF(.NOT. ($000(I) .LT. .10*NT))GO TO 45
 50     CONTINUE
      GO TO 8
 45   STAR(8) = STR
C
C # OF HUNDREDS LT 10% IN Y1 & Y2
C
 8    DO 60 I=1,2
        IF(.NOT. ($100(I) .LT. .10*NT))GO TO 55
 60     CONTINUE
      GO TO 9
 55   STAR(9)=STR
C
C SUM OF ZEROES & HUNDREDS LT 15% IN Y1 & Y2
C
 9    DO 70 I = 1,2
 70     SUMZAH(I) = $000(I) + $100(I)
      DO 80 I = 1,2
        IF ( SUMZAH(I) .GE. .15*NT) GO TO 75
 80     CONTINUE
      GO TO 10
 75   STAR(10)=STR
 10   DO 77  I = 1,TESTS
        IF (STAR(I) .EQ. STR) ADMIT=.FALSE.
 77     CONTINU
      IF(ADMIT) RTURN
      TERMRN = TERMRN + 1.
      IF (TERMRN .GT. 5. .AND. TERMRN .GT. .05*TRYS ) GO TO 999
C
C BYPASS ANALYSIS,PRINT NO.NOK & ESTATS FOR DATASET, GET NEXT DATASET.
C
      RETURN
 999  CONTINUE
C
C COMPUTE & PRINT POP STATS,READ NEXT SET OF CONTROL CARDS.
```

```
C
      NWRITE=0
      PAGE = PAGE + 1
703   WRITE(PRINT,66)TRYS,PAGE
 66   FORMAT('1THIS RUN HAS MORE THAN 5% REJECTED DATA SETS AFTER ',F6.0
     @ ,' TRIES.',T125,'PAGE',I5)
      NWRITE=NWRITE+1
      IF(.NOT.(NWRITE .LT. 2 .AND. NDISC.GT.0))GO TO 704
      PRINT = PRINT + NDISC
      GO TO 703
704   PRINT=PRINT-NDISC
      RMAX = TRYS - TERMRN
      IF (TRYS - TERMRN .GT. 1. - 1E-5) CALL POPS
      RETURN 2
      END
```

```
      SUBROUTINE TRUBLU
     @           (RHO,TM,              NSAMP,SAMPNO,TRUEFF,NP1ST,
     @AGEASS,NESTIM,       GNAT,GTRU,GIES,SCORES,SAMSIZ,
     @ PAGE,GROUP,VARNAM,SEED,NSEED,RCW,NATEFF,ESTMAT,DTASET,NDISC)
      INTEGER*2 M,     AGEORD(300),KOUNT(6),CLAGE(300),CLPRE(300),
     @NCTAL(3,3),RUNS,SAMPNO,        AT(300),STAR(10),DTASET
      LOGICAL*1 ADMIT,PNTOUT,SCORES
      REAL VARNAM(2,5),GROUP(2,3)
      INTEGER PRINT,SEED,SAMSIZ(1)
      DOUBLE PRECISION MEAN(5,3),SD(5,3),COVCOR(5,5,3)
      REAL DATMAT(300,5),ESTMAT(16),
     @GTRU(1),NATEFF,                        Y1(1),Y2(1),AGETM1(1),
     @CUTPRE(6),CUTAGE(3),POSTP(6),POSTC(6),TM(2),NPWT(3,3),POSTTC(3,3),
     @POSTTP(3,3),AGEASS(300) ,IES
      EQUIVALENCE                            (DATMAT(601),Y1(1)),(DATMAT(9
     @01),Y2(1)),(DATMAT(1201),AGETM1(1))
      EQUIVALENCE(POSTP(1),POSTTP(1)),(KOUNT(1),POSTTP(7)),
     @(POSTC(1),POSTTC(1))
      COMMON /LOVERG/ LORUN,VRGEAC
      COMMON /STATS/ MEAN,SD,COVCOR,NT,NC,NP,DATMAT
     @    ,POPMN(5,3),POPSD(5,3),POPCR(5,5,3),AT
      COMMON /BSPREC/BIAS(16,5),PRECIS(16,5),VRTHAT(16,5)
      COMMON/KNOCK/RMAX,TERMRN,TRYS,SEX3,DIFCRS,RATVAR,IES,$000(2),
     @$100(2),SUMZAH(2),PRINT,STAR,ADMIT,PNTOUT
      M = 0
      RUNS = TRYS - TERMRN + 1E-5
      DO 105 J = 1,NESTIM
  105   ESTMAT(J) = 0.
C
C POSTTEST ONLY
C
      M = M+1
      ESTMAT(M) = MEAN(4,3) - MEAN(4,2)
C
C RAW GAINS
C
      M = M+1
      ESTMAT(M) = MEAN(4,3) - MEAN(3,3) - MEAN(4,2) + MEAN(3,2)
C
C STANDARDIZED GAINS
C
      M = M+1
      ESTMAT(M) = MEAN(4,3) - MEAN(4,2) -DSQRT(((NP-1.)*COVCOR(4,4,3)
     @ + (NC-1.)*COVCOR(4,4,2)) / ((NP-1.)*COVCOR(3,3,3) + (NC-1.)*
     @ COVCOR(3,3,2)))*(MEAN(3,3) - MEAN(3,2))
C
C RESIDUAL ANALYSIS ON PRETEST
C
      M = M+1
      ESTMAT(M) = (MEAN(4,3) - MEAN(4,2)) - (COVCOR(4,3,1)/COVCOR(3,3,1)
     @ )*(MEAN(3,3) - MEAN(3,2))
```

```
C
C ANCOVA ON PRETEST
C
      M = M+1
      ESTMAT(M) =  MEAN(4,3)-MEAN(4,2)-(((NP-1.)*COVCOR(4,3,3) + (NC-1.)
     @ *COVCOR(4,3,2)) / ((NP-1.)*COVCOR(3,3,3) + (NC-1.)*COVCOR(3,3,2))
     @ /1.0) * (MEAN(3,3) - MEAN(3,2))
C
C RHO ANCOVA ON PRETEST
C
      M = M+1
      ESTMAT(M) =  MEAN(4,3)-MEAN(4,2)-(((NP-1.)*COVCOR(4,3,3) + (NC-1.)
     @ *COVCOR(4,3,2)) / ((NP-1.)*COVCOR(3,3,3) + (NC-1.)*COVCOR(3,3,2))
     @ /RHO) * (MEAN(3,3) - MEAN(3,2))
C
C BELSEN ANCOVA ON PRETEST
C
      M = M+1
      ESTMAT(M) = MEAN(4,3) - MEAN(4,2) - ((COVCOR(4,3,2)/COVCOR(3,3,2)
     @ /1.0) * (MEAN(3,3) - MEAN(3,2)))
C
C BELSEN & RHO ANCOVA ON PRETEST
C
      M = M+1
      ESTMAT(M) = MEAN(4,3) - MEAN(4,2) - ((COVCOR(4,3,2)/COVCOR(3,3,2)
     @ /RHO) * (MEAN(3,3) - MEAN(3,2)))
C
C SUBCLASSIFICATION ON PRETEST
C
      NSUB  = NP/6
      DO 1  J = 1,5
 1      CUTPRE(J) = (Y1(AT(NC + J*NSUB)) + Y1(AT(NC + J*NSUB + 1)))/2.
      CUTPRE(6) = 0.
      DO 2 J = NP1ST,NT
 2      CLPRE(AT(J)) = (J - NP1ST)/NSUB + 1
      IBOUND = 1
      DO 3 L=1,6
        DO 4 J = IBOUND,NC
          IF ( Y1(AT(J)).LT. CUTPRE(L)) GO TO 5
 4        CLPRE(AT(J)) =  L
 5      IBOUND = J
 3      CONTINUE
      DO 8 I = 1,6
        II = I - 1
        POSTP(I) = 0.
        DO 9 J = 1,NSUB
 9        POSTP(I) = POSTP(I) + Y2(AT(NC + II*NSUB + J))
 8      POSTP(I) = POSTP(I) /  NSUB
      IBOUND = 1
      DO 10 L = 1,6
        POSTC(L) = 0.
        KOUNT(L) = 0
        DO 11 J = IBOUND,NC
          IF (CLPRE(AT(J)) .NE. L) GO TO 12
          POSTC(L) = POSTC(L) + Y2(AT(J))
 11     KOUNT(L) = KOUNT(L) + 1
 12   IF (KOUNT(L) .GT. 0) POSTC(L) = POSTC(L)/KOUNT(L)
        IBOUND = J
 10     CONTINUE
      M = M+1
      ESTMAT(M) = 0.
      J = 0
      DO 13 L = 1,6
        IF (KOUNT(L)) 13,13,16
 16     ESTMAT(M) = ESTMAT(M) + POSTP(L) - POSTC(L)
        J= J + 1
```

```
 13     CONTINUE
        ESTMAT(M) = ESTMAT(M) / J
C
C VALUE ADDED
C
        M = M+1
        ESTMAT(M) = MEAN(4,3) - MEAN(3,3) - (COVCOR(5,3,1)/COVCOR(5,5,1))
       @ *(TM(2) - TM(1))
C
C RESIDUAL ANALYSIS ON PRETEST & AGE
C
        BPRE = (COVCOR(3,4,1) - COVCOR(3,5,1)*COVCOR(4,5,1)) /
       @       (1. - COVCOR(3,5,1)**2)
        BAGE = (COVCOR(4,5,1) - COVCOR(3,5,1)*COVCOR(3,4,1)) /
       @       (1. - COVCOR(3,5,1)**2)
        BPRE = BPRE*SD(4,1)/SD(3,1)
        BAGE = BAGE*SD(4,1)/SD(5,1)
        M = M+1
        ESTMAT(M) = (MEAN(4,3) - MEAN(4,2)) - BPRE*(MEAN(3,3) - MEAN(3,2))
       @ -BAGE*(MEAN(5,3) - MEAN(5,2))
C
C ANCOVA ON PRETEST & AGE
C
        BPRE = ((((NP-1.)*COVCOR(4,3,3) + (NC-1.)*COVCOR(4,3,2))*((NP-1.)
       @ *COVCOR(5,5,3) + (NC-1.)*COVCOR(5,5,2))) - (((NP-1.)*COVCOR(5,4,3
       @ ) + (NC-1.)*COVCOR(5,4,2))*((NP-1.)*COVCOR(5,3,3) + (NC-1.)*
       @ COVCOR(5,3,2)))) / ((((NP-1.)*COVCOR(3,3,3) + (NC-1.)*COVCOR(3,3,
       @ 2)) * ((NP-1.)*COVCOR(5,5,3) + (NC-1.)*COVCOR(5,5,2))) - ((NP-1.)
       @ * COVCOR(5,3,3) + (NC-1.)*COVCOR(5,3,2))**2)
        BAGE =((((NP-1.)*COVCOR(5,4,3) + (NC-1.)*COVCOR(5,4,2))*((NP-1.)*
       @ COVCOR(3,3,3) + (NC-1.)*COVCOR(3,3,2))) - (((NP-1.)*COVCOR(4,3,3)
       @ + (NC-1.)*COVCOR(4,3,2))  * ((NP-1.)*COVCOR(5,3,3)+(NC-1.)*
       @ COVCOR(5,3,2)))) /((((NP-1.)*COVCOR(3,3,3) + (NC-1.)*COVCOR(3,3,
       @ 2)) * ((NP-1.) * COVCOR(5,5,3) + (NC-1.)*COVCOR(5,5,2))) -((NP-1.
       @ )*COVCOR(5,3,3) + (NC-1.)*COVCOR(5,3,2))**2)
        M = M+1
        ESTMAT( M) = (MEAN(4,3)-MEAN(4,2)) - BPRE*(MEAN(3,3)-MEAN(3,2))
       @ -BAGE*(MEAN(5,3) - MEAN(5,2))
C
C RHO ANCOVA ON PRETEST & AGE
C
        BPRE = ((((NP-1.)*COVCOR(4,3,3) + (NC-1.)*COVCOR(4,3,2))*((NP-1.)
       @ *COVCOR(5,5,3) + (NC-1.)*COVCOR(5,5,2))) - (((NP-1.)*COVCOR(5,4,3
       @ ) + (NC-1.)*COVCOR(5,4,2))*((NP-1.)*COVCOR(5,3,3) + (NC-1.)*
       @ COVCOR(5,3,2)))) / (( RHO*
       @                       ((NP-1.)*COVCOR(3,3,3) + (NC-1.)*COVCOR(3,3,
       @ 2)) * ((NP-1.)*COVCOR(5,5,3) + (NC-1.)*COVCOR(5,5,2))) - ((NP-1.)
       @ * COVCOR(5,3,3) + (NC-1.)*COVCOR(5,3,2))**2)
        BAGE =((((NP-1.)*COVCOR(5,4,3) + (NC-1.)*COVCOR(5,4,2))*
       @RHO*((NP-1.)*
       @ COVCOR(3,3,3) + (NC-1.)*COVCOR(3,3,2))) - (((NP-1.)*COVCOR(4,3,3)
       @ + (NC-1.)*COVCOR(4,3,2))  * ((NP-1.)*COVCOR(5,3,3)+(NC-1.)*
       @ COVCOR(5,3,2)))) /(( RHO*
       @                      ((NP-1.)*COVCOR(3,3,3) + (NC-1.)*COVCOR(3,3,
       @ 2)) * ((NP-1.) * COVCOR(5,5,3) + (NC-1.)*COVCOR(5,5,2))) -((NP-1.
       @ )*COVCOR(5,3,3) + (NC-1.)*COVCOR(5,3,2))**2)
        M = M+1
        ESTMAT( M) = (MEAN(4,3)-MEAN(4,2)) - BPRE*(MEAN(3,3)-MEAN(3,2))
       @ -BAGE*(MEAN(5,3) - MEAN(5,2))
C
C BELSEN ANCOVA ON PRETEST & AGE
C
        BPRE=(COVCOR(4,3,2)*COVCOR(5,5,2)-COVCOR(5,4,2)*COVCOR(5,3,2))/
       @(COVCOR(3,3,2)*COVCOR(5,5,2)-COVCOR(5,3,2)**2)
        BAGE=(COVCOR(5,4,2)*COVCOR(3,3,2)-COVCOR(4,3,2)*COVCOR(5,3,2))/
       @(COVCOR(3,3,2)*COVCOR(5,5,2)-COVCOR(5,3,2)**2)
```

```
      M = M+1
      ESTMAT( M) = (MEAN(4,3)-MEAN(4,2)) - BPRE*(MEAN(3,3)-MEAN(3,2))
     @ -BAGE*(MEAN(5,3) - MEAN(5,2))
C
C RHO & BELSEN ANCOVA ON PRETEST & AGE
C
      BPRE=(COVCOR(4,3,2)*COVCOR(5,5,2)-COVCOR(5,4,2)*COVCOR(5,3,2))/
     @(RHO*
     @ COVCOR(3,3,2)*COVCOR(5,5,2)-COVCOR(5,3,2)**2)
      BAGE=(COVCOR(5,4,2)*(RHO*COVCOR(3,3,2))
     @                                -COVCOR(4,3,2)*COVCOR(5,3,2))/
     @(RHO*
     @ COVCOR(3,3,2)*COVCOR(5,5,2)-COVCOR(5,3,2)**2)
      M = M+1
      ESTMAT( M) = (MEAN(4,3)-MEAN(4,2)) - BPRE*(MEAN(3,3)-MEAN(3,2))
     @ - BAGE*(MEAN(5,3) - MEAN(5,2))
C
C
C PRETEST(3) X AGE(3) SUBCLASSIFICATION
C
      NSUB = NP/3
      DO 17 J=1,2
 17     CUTPRE(J) = (Y1(AT(NC + J*NSUB)) + Y1(AT(NC + J*NSUB + 1)))/2.
      CUTPRE(3) = 0.
      DO 18 J=NP1ST,NT
 18     CLPRE(AT(J)) = (J - NP1ST)/NSUB + 1
      IBOUND = 1
      DO 19 L = 1,3
        DO 20 J = IBOUND,NC
          IF (Y1(AT(J)).LT. CUTPRE(L)) GO TO 21
 20       CLPRE(AT(J)) =  L
 21     IBOUND = J
 19     CONTINUE
      DO 22 I = 1,NT
        J = AT(I)
        AGEORD(J) = I
        IF (NC - 1 ) 23,24,24
 23     AGEASS(J) = AGETM1(J)
        GO TO 22
 24     AGEASS(J) = AGETM1(J) + 101.
 22     CONTINUE
      CALL FIZZ (AGEASS,AGEORD,NT)
      IF (.NOT.SCORES) GO TO 60
 80   FORMAT (1HO,T60,'AGE IN PRETEST ORDER',I2)
 81   FORMAT(1H , 10F13.3)
      WRITE(PRINT,80) NSAMP
      WRITE(PRINT,81) (AGETM1(AT(J)),J=1,NT)
 60   DO 25 J = 1,2
 25     CUTAGE(J) = (AGETM1(AGEORD(NC + J*NSUB)) + AGETM1(AGEORD(NC +
     @  J*NSUB+ 1)))/2.
      CUTAGE(3) = 0.
      DO 26 J = NP1ST,NT
 26     CLAGE(AGEORD(J)) = (J - NP1ST)/NSUB + 1
      IBOUND = 1
      DO 27 L = 1,3
        DO 28 J = IBOUND,NC
          IF ( AGETM1(AGEORD(J)).LT. CUTAGE(L)) GO TO 29
 28       CLAGE(AGEORD(J)) = L
 29     IBOUND = J
 27     CONTINUE
      DO 31 I = 1,3
        DO 31 J = 1,3
          NCTAL(I,J) = 0
          NPWT(I,J) = 0.
          POSTTC(I,J) = 0.
 31       POSTTP(I,J) = 0.
```

```
      DO 32 I = 1,NT
        J = AT(I)
        IF (I - NC) 33,33,34
33      L = CLPRE(J)
        K = CLAGE(J)
        POSTTC(L,K) = POSTTC(L,K) + Y2(J)
        NCTAL(L,K) = NCTAL(L,K) + 1
        GO TO 32
34      L = CLPRE(J)
        K = CLAGE(J)
        POSTTP(L,K) = POSTTP(L,K) + Y2(J)
        NPWT(L,K) = NPWT(L,K) + 1.
32      CONTINUE
      DO 36 I = 1,3
        DO 36 J = 1,3
          IF (NCTAL(I,J) .GT. 0) POSTTC(I,J) = POSTTC(I,J)/NCTAL(I,J)
36        CONTINUE
      DO 37 I = 1,3
        DO 37 J = 1,3
          IF (NPWT(I,J) .GT. 0.) POSTTP(I,J) = POSTTP(I,J)/NPWT(I,J)
37        CONTINUE
      M = M+1
      DO 38 I = 1,3
        DO 38 J = 1,3
          IF (NCTAL(I,J) .EQ. 0.OR.NPWT(I,J).EQ.0.)GO TO 38
          NPWT(I,J) = NPWT(I,J) / NP
          ESTMAT( M) = ESTMAT( M) + NPWT(I,J)*(POSTTP(I,J)-POSTTC(I,J))
38        CONTINUE
      IF (PNTOUT) CALL PRINTR(PAGE,DTASET,          ROW,TRUEFF,NATEFF,
     %     NSAMP,ESTMAT,VARNAM,GROUP,SEED,NSEED,NESTIM,&77,NDISC)
77    DO 40 I = 1,NESTIM
      PRECIS(I,NSAMP) = PRECIS(I,NSAMP) + ESTMAT(I)**2
40      BIAS(I,NSAMP) = BIAS(I,NSAMP) + ESTMAT(I)
      GTRU(NSAMP) = GTRU(NSAMP) + TRUEFF
      IF (NSAMP .GT. 1) GO TO 74
C
C RUNNING GRAND TOTALS GNAT,GIES,GTRU(NSAMP), I=1,SAMPNO
C
      GNAT = GNAT + NATEFF
      GIES = GIES + IES
      DO 61 K = 1,3
        DO 61 J = 1,5
          POPMN(J,K)=POPMN(J,K)+MEAN(J,K)
          POPSD(J,K)=POPSD(J,K) + SD(J,K)
          DO 61 I = 1,5
61          POPCR(I,J,K) = POPCR(I,J,K)+CORCOR(I,J,K)
74    IF (RUNS .GE. RMAX - 1E-5 .AND. NSAMP .EQ. SAMPNO) GO TO 41
C
C MAXRUN CALLED ONLY ONCE PER SIMULATION.
C
       IF (RUNS .EQ. LORUN      .AND. NSAMP .EQ. 1)CALL MAXRUN(NESTIM,
     @NBIAS,PRINT,RMAX,PAGE,DTASET,NDISC)
      RETURN
      ENTRY POPS
41    NSAMP = 1
      GNAT = GNAT/RUNS
      GIES = GIES/RUNS
      DO 73 K = 1,3
        DO 73 J = 1,5
          POPMN(J,K) = POPMN(J,K) / RUNS
          POPSD(J,K) = POPSD(J,K) /RUNS
          DO 73 I = 1,5
73          POPCR(I,J,K) = POPCR(I,J,K)/RUNS
72    GTRU(NSAMP) = GTRU(NSAMP) / RUNS
      DO 50 I = 1,NESTIM
        PRECIS(I,NSAMP)=(PRECIS(I,NSAMP)-2*GTRU(NSAMP)*BIAS(I,NSAMP)
```

```
     Z )/RUNS + GTRU(NSAMP)**2
       BIAS(I,NSAMP) = BIAS(I,NSAMP) / RUNS  - GTRU(NSAMP)
      VRTHAT(I,NSAMP) = PRECIS(I,NSAMP) - BIAS(I,NSAMP)**2
50     CONTINUE
      CALL TRAIL (NSAMP,RUNS,PAGE,GNAT,GTRU(1),GIES,SAMSIZ,
     2SAMPNO,NESTIM,NDISC,ROW,VARNAM,GROUP,SEED,NSEED)
      NSAMP = NSAMP + 1
      IF (NSAMP .LE. SAMPNO) GO TO 72
      RETURN
      END




      SUBROUTINE MAXRUN (NESTIM,NBIAS,PRINT,RMAX,PAGE,DTASET,NDISC)
57    FORMAT(' ESSD',16F8.4)
59    FORMAT(5X,16I8)
15    FORMAT(1H1,123(1H=),'PAGE ',I4)
16    FORMAT(1HO,'N TO CONVERGE = ',F8.0)
      INTEGER PRINT
      INTEGER*2 PAGE,DTASET
      COMMON /BSPREC/BIAS(16,5),PRECIS(16,5),VRTHAT(16,5)
      COMMON /LOVERG/ LORUN,VRGFAC
      REAL ESMEAN(16),ESTSD(16)
      EQUIVALENCE(ESMEAN(1),BIAS(1,1)),(ESTSD(1),PRECIS(1,1))
      DO 1  L=1,NESTIM
        ESMEAN(L) = ESMEAN(L) / LORUN
1       ESTSD(L) = VRGFAC*(ESTSD(L)/LORUN - ESMEAN(L)**2)
      RVERGE = ESTSD(1)
      DO 4 L = 2,NESTIM
        IF (ESTSD(L) .GT. RVERGE) RVERGE = ESTSD(L)
4       CONTINUE
      IF (RVERGE - RMAX) 5,7,7
5     IF (RVERGE - LORUN) 6,6,8
8     RMAX = RVERGE
      GO TO 7
6     RMAX = LORUN + 1.
7     DO 9 L = 1,NESTIM
9       ESTSD(L) = SQRT(ESTSD(L) / VRGFAC )
      PAGE = PAGE + 1
      PRINT = PRINT + NDISC
      WRITE (PRINT,15) PAGE
      WRITE(PRINT,16)RVERGE
      WRITE (PRINT,59) (I,I=1,NESTIM)
      WRITE(PRINT,57)ESTSD
      PRINT = PRINT - NDISC
      DO 10 L = 1,NESTIM
        PRECIS(L,1) = LORUN*(ESTSD(L)**2+ESMEAN(L)**2)
10      BIAS(L,1) = LORUN*ESMEAN(L)
      DTASET=3
      RETURN
      END
```

```
      SUBROUTINE PRINTR(PAGE,DTASET,            POW,TRUEFF,NATEFF,
     @    NSAMP,ESTMAT,VARNAM,GROUP,SEED,NSEED,NESTIM,*,NDISC)
      LOGICAL*1 ADMIT,PNTOUT
      INTEGER*2 DTASET,PAGE,      RUNS ,STAR(10),AT(300),
     @SAMPNO
      INTEGER PRINT,NGROUP(3),ROW,STATIC(3,5),SEED,BRIEF(4),SAMSIZ(1)
      COMMON /BSPREC/BIAS(16,5),PRECIS(16,5),VPTHAT(16,5)
      REAL DATMAT(300,5),IES,NATEFF,ESTMAT(16),VARNAM(2,5),GROUP(2,3)
      DOUBLE PRECISION COVCOR(5,5,3),MEANS(5,3),SDS(5,3)
      DATA              STATIC/'MEAN','S',' ','POP ','MEAN','S','POP ',
     @'SD''','S','COVA','RIAN','CES','CORR','FLAT','ICNS'/,BRIEF/'THAT',
     @'BIAS','V(T)','PREC'/
      EQUIVALENCE (NT,NGROUP(1)),(NC,NGROUP(2)),(NP,NGROUP(3))
      COMMON/KNOCK/RMAX,TERMRN,TRYS,SEX3,DIFCRS,RATVAR,IES,$000(2),
     @$100(2),SUMZAP(2),PRINT,STAR,ADMIT,PNTOUT
      COMMON/STATS/MEANS,SDS,COVCOR,NT,NC,NP,DATMAT,
     @    POPMN(5,3),POPSD(5,3),POPCR(5,5,3),AT
      DTASET = DTASET + 1
      PRINT = PRINT + NDISC
      IF(ADMIT) GO TO 725
      PRINT=PRINT-NDISC
      NWRITE = 0
      IF(.NOT.ADMIT .AND. DTASET .LT. 4 .AND. NSAMP .EQ. 1)GO TO 727
725   IF (DTASET .LT. 4) GO TO 1
      DTASET = 1
      PAGE = PAGE + 1
727   WRITE(PRINT,100) TRYS,NSAMP,((GROUP(I,J),I=1,2),NGROUP(J),J=1,3),
     @ PAGE
      GO TO 2
 1     WRITE (PRINT,101) TRYS,NSAMP
     @                  ,((GROUP(I,J),I=1,2),NGROUP(J),J=1,3)
 2    WRITE (PRINT,102) (VARNAM,I=1,3)
      WRITE (PRINT,103) (STATIC(I,1),I=1,3),MEANS
      WRITE (PRINT,104)(STATIC(I,4),I=1,3),SEED,(STATIC(I,4),I=1,3),
     @NSEED,(STATIC(I,4),I=1,3)
      DO 3 I = 1,ROW
      WRITE(PRINT,105) (VARNAM(J,I),J=1,2),(COVCOR(I,J,1),J=1,I)
      WRITE (PRINT,106) (COVCOR(I,J,2),J=1,I)
   3  WRITE (PRINT,107) (COVCOR(I,J,3),J=1,I)
      WRITE (PRINT,108)((STATIC(I,5),I=1,3),J=1,3)
      DO 4 J = 2,ROW
      JJ = J - 1
      WRITE (PRINT,105) (VARNAM(I,J),I=1,2),(COVCOR(I,J,1),I=1,JJ)
      WRITE (PRINT,106) (COVCOR(I,J,2),I=1,JJ)
 4    WRITE (PRINT,107) (COVCOR(I,J,3),I=1,JJ)
      IF(NSAMP .GT. 1 .AND. .NOT. ADMIT)GO TO 711
      GO TO 715
711   NWRITE=NWRITE+1
      IF (.NOT.(NWRITE .LT. 2 .AND. NDISC .GT. 0)) GO TO 726
      GO TO 727
```

```
715   IF(NSAMP .GT. 1) GO TO 20
      IF (ADMIT) WRITE (PRINT,113)
      IF (.NOT.ADMIT) WRITE(PRINT,114)
      WRITE(PRINT,11) TRYS,SEX3,DIFORS,STAR(1),RATVAR,STAR(2),COVCOR
     @ (3,4,3),STAR(3),COVCOR(3,4,2),STAR(4),
     @ IES,STAR(5),(COVCOR(I,5,3),I=1,4),STAR(6),(COVCOR(I,5,2),I=1,4),
     @ STAR(7),$900,STAR(8),$100,STAR(9),SUMZAH,STAR(10)
      IF(ADMIT)GO TO 20
      NWRITE = NWRITE + 1
      IF(NWRITE .LT. 2) PRINT=PRINT+NDISC
      IF(.NOT.(NWRITE .LT. 2 .AND. NDISC .GT. 0)) GO TO 726
      GO TO 727
726   PRINT = PRINT - NDISC
      IF (.NOT.ADMIT)RETURN 1
 20   WRITE (PRINT,12 ) TRUEFF,NATEFF,IES
      WRITE (PRINT,59) (I,I=1,NESTIM)
      WRITE (PRINT,67) BRIEF(1),ESTMAT
      PRINT = PRINT -NDISC
      RETURN
      ENTRY TRAIL(NSAMP,RUNS,PAGE,              GNAT,GTRU1,GIES,SAMSIZ,
     @SAMPNO,NESTIM,NDISC,ROW,VARNAM,GROUP,SEED,NSEED)
      NWRITE = 0
 72   IF (NSAMP .GT. 1) GO TO 33
      PAGE = PAGE + 1
      WRITE (PRINT,51) RUNS,PAGE
      WRITE (PRINT,13) GROUP
      WRITE (PRINT,102) (VARNAM,I=1,3)
      WRITE (PRINT,103) (STATIC(I,2),I=1,3),PCPMN
      WRITE (PRINT,103) (STATIC(I,3),I=1,3),PCPSD
      WRITE (PRINT,104)(STATIC(I,4),I=1,3),SEED,(STATIC(I,4),I=1,3),
     @NSEED,(STATIC(I,4),I=1,3)
      DO 8 I = 1,ROW
      WRITE(PRINT,105) (VARNAM(J,I),J=1,2),(PCPCR (I,J,1),J=1,I)
      WRITE (PRINT,106) (PCPCR (I,J,2),J=1,I)
 8    WRITE (PRINT,107) (PCPCR (I,J,3),J=1,I)
      WRITE (PRINT,108)((STATIC(I,5),I=1,3),J=1,3)
      DO 9 J = 2,ROW
      JJ = J - 1
      WRITE (PRINT,105) (VARNAM(I,J),I=1,2),(PCPCR (I,J,1),I=1,JJ)
      WRITE (PRINT,106) (PCPCR (I,J,2),I=1,JJ)
 9    WRITE (PRINT,107) (PCPCR (I,J,3),I=1,JJ)
      WRITE (PRINT,58) GTRU1,GNAT,GIES
 58   FORMAT(' ',T31,'GRAND TRUE =',F8.2,3X,'GRAND NAT =',F8.2,3X,
     # 'GRAND IES =',F8.2,'%')
 33   WRITE (PRINT,56) NSAMP,SAMSIZ(NSAMP ),(J,J=2,NESTIM)
      WRITE (PRINT,57)BRIEF(2),(BIAS(I,NSAMP),I=1,NESTIM)
      WRITE (PRINT,57) BRIEF(3),(VRTHAT(I,NSAMP),I=1,NESTIM)
      WRITE (PRINT,57) BRIEF(4),(PRECIS(I,NSAMP),I=1,NESTIM)
      IF (NSAMP .EQ. SAMPNO)WRITE(PRINT,55)(I,I=1,NESTIM)
      NWRITE = NWRITE + 1
      IF(NWRITE .LT. 2) PRINT=PRINT+NDISC
      IF(.NOT.(NWRITE .LT.2 .AND. NDISC .GT. 0)) GO TO 73
      GO TO 72
 73   PRINT = PRINT-NDISC
      RETURN
```

```
59    FORMAT(5X,16I8)
100   FORMAT(1H1,'RUN ',F5.0,' SAMPLE',I2,T29,2A4,'=',I3,T69,2A4,
     @'=',I3,T109,2A4,'=',I3,T125,'PAGE ',I4)
101   FORMAT(1H ,'RUN ',F5.0,' SAMPLE',I2,T29,2A4,'=',I3,T69,2A4,
     @'=',I3,T109,2A4,'=',I3)
102   FORMAT(T14,30A4)
103   FORMAT(1H ,3A4,15F8.3)
104   FORMAT(1H ,T29,3A4,2X,'SEED IS',I13,T69,3A4,2X,'SEED # IS',
     @I13,T109,3A4)
105   FORMAT(1H ,2A4,T14,5F8.3)
106   FORMAT (1H+,T54,5F8.3)
107   FORMAT (1H+,T94,5F8.3)
108   FORMAT(1H ,T14,3(15X,3A4,13X))
13    FORMAT(1H ,T29,2A4,T69,2A4,T109,2A4)
12    FORMAT(T29,'T RATS FOR DIFFERENT ANALYSES. TRUE EFFECT =',F8.3,
     @',  NATURAL EFFECT =',F8.3,',  IES =',F7.2,'%')
111   FORMAT(1H ,T14,(I8))
113   FORMAT(1H ,'ACCEPTED:')
114   FORMAT(1H ,'REJECTED:')
11    FORMAT(1H+,T12,F6.0
     @                    ,':',3X,'5*SE =',F8.3,' AND |Y(1P.)-Y(1C.)
     @| =',F8.3,A1,3X,'VAR(Y(1P.))/VAR(Y(1C.)) =',F6.2,A1,3X,'R(PRE P,PO
     @ST P) =',F6.4,A1  /4X,'R(PRE C,POST C) =',F6.4,A1,3X,'IES =',F8.2,
     @'%',A1,3X,'AGE CORRELATIONS W/I P =',4F7.4,A1    /
     @4X,'AGE CORRELATIONS W/I C =',4F7.4,A1,3X,'# OF ZEROES FOR Y1 & Y2
     @ =',2F5.0,A1   /  4X,'# OF HUNDREDS FOR Y1 & Y2 =',2F5.0,A1,3X,
     @ 'SUM OF 0''S & 100''S FOR Y1 & Y2 =',2F5.0,A1)
51    FORMAT(1H1,T56,'SUMMARY AFTER'
     @,I4,' RUNS.',T125,'PAGE',I5)
56    FORMAT(' SAMPLE',I2,T13,'1N=',I3/'+',T14,15I8)
57    FORMAT(1H ,A4,16F8.3 )
67    FORMAT(1H ,A4,16F8.3/)
55    FORMAT(1H0,'THE LINEUP'      /1H+,10(1H_)/1H0,I5,'. POSTTEST ONLY'/
     11H ,I5,'. RAW GAINS'/1H ,I5,'. STANDARDIZED GAINS'/1H ,I5,'. RESID
     2UAL ANALYSIS ON PRETEST'/1H ,I5,'. ANCOVA ON PRETEST'/1H ,I5,'. RH
     30 ANCOVA ON PRETEST'/1H ,I5,'. BELSEN ANCOVA ON PRETEST'/1H ,
                                                              I5,'. B
     4ELSEN & RHO ANCOVA ON PRETEST'/1H ,I5,'. SUBCLASSIFICATION ON PRET
     5EST(6)'/1H ,I5,'. VALUE ADDED'              /1H ,I5,'. RESIDUAL ANAL
     6YSIS ON PRETEST & AGE'/1H ,I5,'. ANCOVA ON PRETEST & AGE'/1H ,I5,
                                                              '.
     &
     6RHO ANCOVA ON PRETEST & AGE'/1H ,I5,'. BELSEN ANCOVA ON PRETEST &
     7AGE'/1H ,I5,'. BELSEN & RHO ANCOVA ON PRETEST & AGE'/1H ,I5,'. PRE
     8TEST(3) BY AGE(3) SUBCLASSIFICATION')
      END
```